

# QUANTILE TOMOGRAPHY: USING QUANTILES WITH MULTIVARIATE DATA

LINGLONG KONG AND IVAN MIZERA

**ABSTRACT.** Directional quantile envelopes—essentially, depth contours—are a possible way to condense the directional quantile information, the information carried by the quantiles of projections. In typical circumstances, they allow for relatively faithful and straightforward retrieval of the directional quantiles, offering a straightforward probabilistic interpretation in terms of the tangent mass at smooth boundary points. They can be viewed as a natural, nonparametric extension of “multivariate quantiles” yielded by fitted multivariate normal distribution, and, as illustrated on data examples, their construction can be adapted to elaborate frameworks—like estimation of extreme quantiles, and directional quantile regression—that require more sophisticated estimation methods than simply evaluating quantiles for empirical distributions. Their estimates are affine equivariant whenever the estimators of directional quantiles are translation and scale equivariant; mathematically, they express the dual aspect of directional quantiles.

## 1. INTRODUCTION

**1.1. Objective.** The article aims at addressing certain aspects of using quantiles to obtain insights about multivariate data.

While such an objective could be mistaken for yet another attempt in the ongoing quest for “multivariate quantiles”—as thoroughly reviewed by Serfling (2002)—we would like to stress that we differ in the position that *no multivariate generalization of the quantile concept may be needed at all*—only a way how to condense and present the information about the quantiles of certain univariate attributes of the data, so that the specific and well-recognizable *meaning* of quantiles, as expounded below, transcribes into the multivariate context.

---

2000 *Mathematics Subject Classification.* Primary 62H05, Secondary 52A07, 62G08.

*Key words and phrases.* Multivariate Quantile Contours, Data Depth, Quantile Regression.

Research supported by the Natural Sciences and Engineering Research Council of Canada.

However, it would be negligent to claim that other contributions to the multivariate quantile topic did not pursue similar objectives too; therefore, our efforts may be viewed in a certain vicinity of theirs.

**1.2. Outline.** Our somewhat unconstructive attitude makes the task more difficult: instead of revealing the essence of our proposal immediately, preferably in the first section—a widely recommended stylistic strategy allowing a busy reader to skip subsequent predictable apology—we must urge the reader to go through the bulk of the following sections, to get our message undistorted. These include, in particular, Section 2, where the quantile concept, but more importantly, its meaning in are reviewed; Section 3, which opens a discussion of the multivariate specifics culminating in the introduction of directional quantiles in Section 4. Sections 5 and 6 then digress into potential antitheses, just to form a Hegel-like dialectic synthesis with our proposed methodology in Section 7.

For really hard-pressed individuals, a possible rapid route—without any warranty, however—could be (beyond browsing the references and acknowledgments in Section 13) to skim over the figures and their captions, and then read the conclusion in Section 12.

On the other hand, an interested reader might like to learn from Section 10 about possible shortcomings of our proposed methodology, and from Section 9 about its outstanding flexibility in adapting to more sophisticated situations, like quantile regression or the estimation of extreme quantiles. Finally, some readers may appreciate that all proofs are given in the Appendix.

As the only spoiler, we reveal that our forthcoming deliberations will bring us very close to what is already known in statistical sciences as halfspace (Tukey) depth; we only hope that the reader would rightly perceive this as an *a posteriori* inevitable outcome rather than *à la these* presumption. In fact, one of the outcomes is that not that much may be halfspace depth relevant to quantiles as quantiles to halfspace depth—whatever its *raison d'être* otherwise may be.

## 2. QUANTILES: A REVIEW

**2.1. Definition and terminology.** The concept of the quantile function of a univariate probability distribution is known well beyond any need of exposition; a casual reference like Shorack (2000) or the encyclopedic entry of Eubank (1986) results in the following.

**Definition 1.** For  $0 < p < 1$ , the  $p$ -th *quantile* of a distribution  $P$  is defined to be

$$Q(p) = \inf\{u: F(u) \geq p\},$$

where  $F(u) = P((-\infty, u])$  is the cumulative distribution function of  $P$ .

The most prominent quantile corresponds to  $p = 1/2$  and is called *median*. Other quantiles do often have traditional names too—for instance, those indexed by  $p = 1/n, 2/n, \dots, (n-1)/n$ , with  $n = 4, 5, 10$  are known as *quartiles*, *quintiles*, and *deciles*, respectively; the relevant linguistic aspects are discussed by Aronson (2001). A preference for percentages in the general populace is reflected by a synonymous term *percentile*, indexed by  $100p$  rather than  $p$ .

**2.2. Ambiguous and void quantiles.** Essentially, we tend to view  $Q$  as a function of  $p$  inverse to  $F$ , that is, a function satisfying

$$(1) \quad F(Q(p)) = p.$$

However, a simplified definition via the identity (1) would work only in regular cases; the sophistication of Definition 1 is needed to handle situations when there is none, or more than one  $Q(p)$  satisfying (1). In this connection, it is useful to invoke the following alternative quantile definition via minimization of the “check” function  ${}_{p-1}|x|_p = x(p - I(x < 0))$ .

**Definition 2.** For  $0 < p < 1$ , the  $p$ -th *quantile set* of a distribution  $P$  is defined to be the set  $\mathcal{Q}(p)$  of all  $q$  minimizing the integral

$$\int {}_{p-1}|x - u|_p P(dx).$$

Every prescription that results in a unique element of every  $p$ -th quantile set—like that given by Definition 1—will be referred to as a *quantile version*.

The definition exploits a well-known fact that every quantile set is closed interval, possibly a singleton; the latter case takes place, in particular, when there is no  $Q(p)$  satisfying (1), the case we refer to as *void* quantile. If, on the other hand, the set of  $q$  such that  $F(q) = p$  is nonempty, then this set is equal to the quantile set; if it contains more than one element, we speak about *ambiguous* quantiles—“multiply realizable quantiles” in the terminology of Shorack (2000).

While working with a set-valued definition may have theoretical advantages, practitioners rather demand suitable quantile version: either the inf, “theoretical” one given by Definition 1, or some other choice. Hyndman and Fan (1996) review those

in practical use, which are also implemented as options of the R function `quantile` (Frohne and Hyndman, 2004), whose documentation may thus serve as a quick overview. For the computations in this article, we used the “`type=1`” version of `quantile`, abiding by the theoretical Definition 1, with the objective to produce maximally faithful illustrations of the explained theoretical facts. In concrete applications we might rather consider one of the interpolated versions—say, the R default.

It is not hard to see that quantile ambiguity can occur only when the distribution contains a “gap”, an open interval such that  $P((a, b)) = 0$ , and quantile voidness if the distribution contains an atom, a point  $c$  such that  $P\{c\} > 0$ . These phenomena can be often ruled out for population distributions—and although inevitable for empirical ones, it should be noted that their extent often vanishes with growing sample size.

**2.3. The meaning of quantiles.** As affirmed by Eubank (1986), and others—we recommend, in particular, Parzen (2004) and references there—quantiles play a fundamental and multifaceted rôle in statistics. Nevertheless, in this article we intentionally ignore all the potential variety and accent only the direct probabilistic interpretations, those coming as an immediate consequence of the definition.

An example may perhaps make the thrust of this intent more clear. Suppose that we just learned the outcome of an examination. If our score is 54, and we know that the 0.9-th quantile of the class distribution is 50, then we know that we can count ourselves among the proud top 10% of achievers. Similarly, if the score is 20, and we know that 0.1-th quantile is 18, then we may not feel that great—until realizing that the class median is 25, which tells us that 40% of the fellow students share a similar mediocre fate.

If we regurgitate such banal facts—the descriptive potential of the quantiles was already pointed out by Quetelet and endorsed by Edgeworth (1886, 1893) and Galton (1888-1889)—we do it with a sole intent: *to ensure that the reader understands what we find important about quantiles*. We neither deny, nor neglect, nor give up on more sophisticated applications; however, we strongly believe that none of these is worth losing the descriptive grip illustrated in our parable above.

### 3. BEYOND MARGINAL VISTAS

**3.1. A bivariate example.** To exemplify how quantiles may be useful with multivariate data, let us introduce a bivariate example. Figure 1 shows the scatterplot of the weight and height of 4291 Nepali children, aged between 3 and 60 months—the

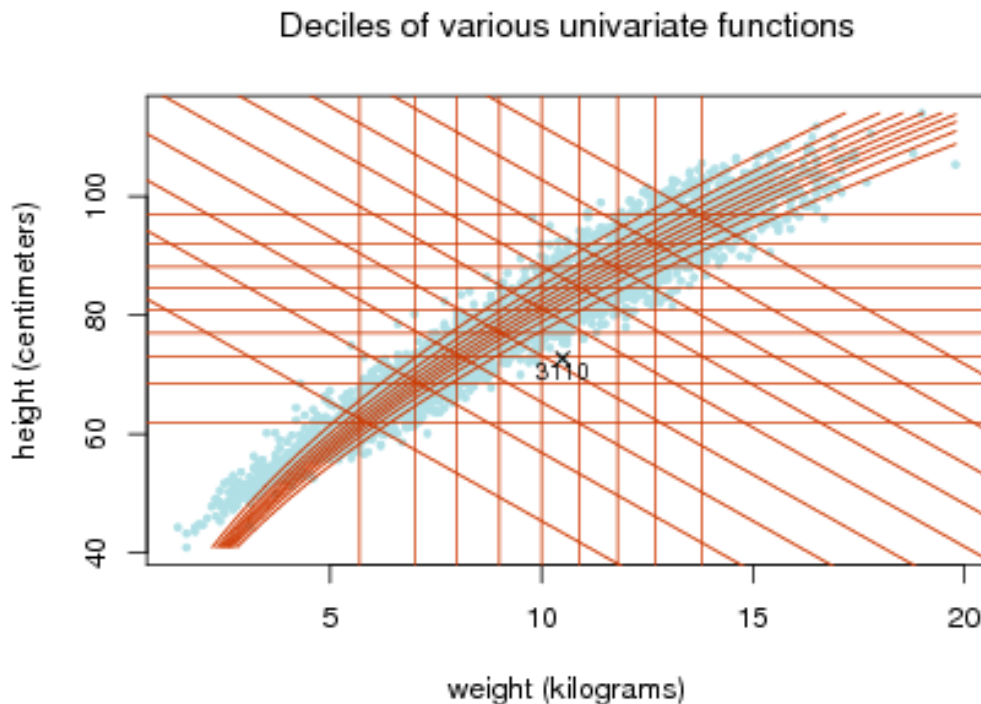


FIGURE 1. *Multivariate data typically offer insights beyond the marginal view, often through the quantiles of univariate functions of primary variables. Plotting the corresponding quantile lines is an appealing way to present this information.*

data constituting a part of the Nepal Nutrition Intervention Project-Sarlahi (NNIP-S, principal investigator Keith P. West, Jr., funded by the Agency of International Development).

The horizontal and vertical lines show the deciles of height and weight, respectively, of the empirical distributions of the corresponding variables. The simple conclusions that can be inferred are akin to those in the univariate case; for instance, we can see that the points above the upper horizontal line correspond to 10% of the subjects exceeding the others in height; similarly, the points right of the rightmost line correspond to 10% of those exceeding the others in weight.

It would be interesting to know what proportion of the data corresponds to the upper right corner, but this information is not directly available (unless we count

the points manually). Also, regarding the subject labeled by 3110, we can only say that its weight is somewhat higher than, but otherwise fairly close to the median; its height is about at the second decile, that is, exceeding about 20% and exceeded by about 80% of its peers.

Nevertheless, the reader will probably agree that Figure 1 indicates that 3110 is in certain sense extremal, outstanding from the rest.

**3.2. Functions of primary variables.** A possible way of substantiating this impression quantitatively is to invoke Quetelet’s *body mass index* (hereafter *BMI*), defined as the ratio of weight to squared height<sup>1</sup>. The curved lines in Figure 1 show the deciles of the empirical distribution of the *BMI*. We can see that in terms of *BMI*, the subject 3110 is indeed extreme, belonging to the group of 10% of those with maximal *BMI*.

An expert on nutrition may dispute the relevance of *BMI* for young children, and remind us of possible alternatives—for instance, the Rohrer index (ratio of weight to *cubed* height, hereafter *ROI*). However, we do not think that the problem lies in deciding whether that or another index is to be preferred; the essence of the data may lie well beyond the index-style of description.

For example, suppose that we would like to make quantitative statements about the subjects represented by the points in the upper right and lower left rectangles. Since we are not aware of any relevant index related to this objective, we may simply look, in Figure 1, at the deciles of some suitable linear combination of weight and height.

Pursuing vague objectives in nonlinear realm may be hard—there are simply too many choices. A possible solution is to limit the attention only to *linear* functions of the original data; note that the “*BMI* contours” in Figure 1 are not that badly approximated by straight lines. In fact, our example offers even a better solution: by taking the *logarithms* of weight and height as primary variables, we will be able to investigate both *BMI* and *ROI* (and possibly much more) among their linear combinations. Therefore, beginning with Figure 2, we use the logarithmic scale.

However, rather with this rather technical detail, we would like to conclude this section more substantively by expressing our opinion that quantiles of certain functions of variables (in particular, linear combinations) may provide a valuable information about multivariate data, and that plotting the corresponding (directional) quantile

---

<sup>1</sup>In metric units; imperial sources may include an adjusting multiplier.

lines is an appealing way to present this information—in particular, to indicate how the quantiles divide the data.

#### 4. DIRECTIONAL QUANTILES

**4.1. The definition of directional quantiles.** Notationally, it is often convenient to work with random variables/vectors, and write

$$Q(p) = Q(p, X) = \inf\{u: \mathbb{P}[X \leq u] \geq p\},$$

despite that the quantiles depend only on the distribution,  $P$ , of  $X$ . (The apparent notational convention hereafter is to suppress the dependence on  $X$  when no confusion may arise.)

Once we focus on linear combinations, we realize that it is sufficient to look exclusively on projections; any other linear combination is a multiple of a projection, and the quantile of a multiple is the multiple of the quantile. The following definition and theorem are elementary, but in a sense fundamental.

**Definition 3.** An operator that assigns a point, or set of points,  $T$ , to a random variable  $X$  is called *translation equivariant*, if its value for  $X + b$  coincides with  $T + b$ , and *scale equivariant*, if its value for  $cX$  coincides with  $cT$ . (If  $T$  is a set, then the transformations are performed elementwise.)

**Theorem 1.** For every  $p \in (0, 1)$ , the quantile operator  $Q(p, \cdot)$  is translation and scale equivariant.

Informally, directional quantiles are quantiles of the projections into the direction of  $s$  and the directional quantile lines are the lines indicating how these quantiles divide the data.

**Definition 4.** We call any vector with unit norm in  $\mathbb{R}^d$  a *normalized direction*, and denote the set of all such vectors by  $\mathbb{S}^{d-1}$ . Let  $X$  be a random vector with distribution  $P$ . Given a normalized direction  $s \in \mathbb{S}^{d-1}$  and  $0 < p < 1$ , the  $p$ -th *directional quantile*, in the direction  $s$ , is defined as the  $p$ -th quantile of the corresponding projection of the distribution of  $X$ ,

$$Q(p, s) = Q(p, s, X) = Q(p, s^T X).$$

The corresponding  $p$ -th *directional quantile hyperplane* (line for  $d = 2$ ) is given by the equation  $s^T x = Q(p, s)$ . The  $p$ -th *directional quantile set* is defined analogously as

$$\mathcal{Q}(p, s) = \mathcal{Q}(p, s, X) = \mathcal{Q}(p, s^T X).$$

The  $p$ -th directional quantile in the direction  $s$  and the  $(1 - p)$ -th directional quantile in the direction  $-s$  are not necessarily equal, due to the inf convention employed in Definition 1. Nonetheless, they often coincide—for instance, it is not possible to distinguish between any  $p$ -th and  $(1 - p)$ -th directional quantile lines if quantile ambiguity is excluded for any projection of  $P$ . A sufficient condition for that is the following property.

**Definition 5.** Let  $P$  be a probability distribution in  $\mathbb{R}^d$ . We say that  $P$  has *contiguous support*, if there is no intersection of halfspaces with parallel boundaries that has nonempty interior but zero probability  $P$  and divides the support of  $P$  to two parts.

Note that if the support is not contiguous, it is not connected; however, it may be disconnected and still contiguous. We believe that the contiguous support, and thus the lack of quantile ambiguity, is a fairly typical virtue of population distributions, and consequently will limit our attention to  $p$  from  $(0, 1/2]$  (although some subsequent theorems will be formulated for more general  $p$ ).

**4.2. Directional quantile information.** Figure 2 shows the plot of the logarithms of weight and height, together with superimposed lines indicating deciles in 20 uniformly spaced directions. While we still champion plotting directional quantile lines as an appealing way to present the directional quantile information, we have to admit that the plot becomes quickly overloaded if multiple directions and indexing probabilities are requested.

Therefore, we would like to consider alternatives aimed at compression of the directional quantile information. While our focus is not exclusively graphical, the task of plotting is probably the most palpable one to epitomize this objective. Before getting to the essence, however, we would like to make two digressions—in a hope that they will help to clarify our stance.

## 5. QUANTILE BILOTS

**5.1. Another way to visualize directional quantiles.** It looks like that the mere task of plotting of directional quantiles—for fixed  $p$ —should not pose any special difficulty: the directional quantile  $Q(p, s)$  may be represented by the point  $Q(p, s)s$  lying on the line with direction  $s$ . Such a plot is more informative when superimposed on the original bivariate scatterplot; this leads to an amalgam which we decided to call a *quantile biplot*, and whose instances can be seen in Figure 3.

In the left panel, we used the same coordinate system for both data and quantile component, the system inherited from the data. The origin thus happens to be located outside the data cloud; as a consequence, the “quantile contour” extends far away from the data cloud, and intersects itself. We can avoid some of these effects by choosing the origin for the quantile plotting inside the data cloud, the possibility shown in the right panel of Figure 3. Finding the appropriate location takes some trial and error; the coordinate-wise median is a decent guess (as indicated also by some theory; see Theorem 10 below).

**5.2. Continuity.** The line of directional quantiles in a quantile biplots appears to be a continuous curve. However, this can be just an artifact of the plotting rather than a rule, and hence deserves some introspection. The following theorem shows that the continuity of directional quantiles, the continuity of  $Q(p, s, X)$  in  $s$ , is quite common.

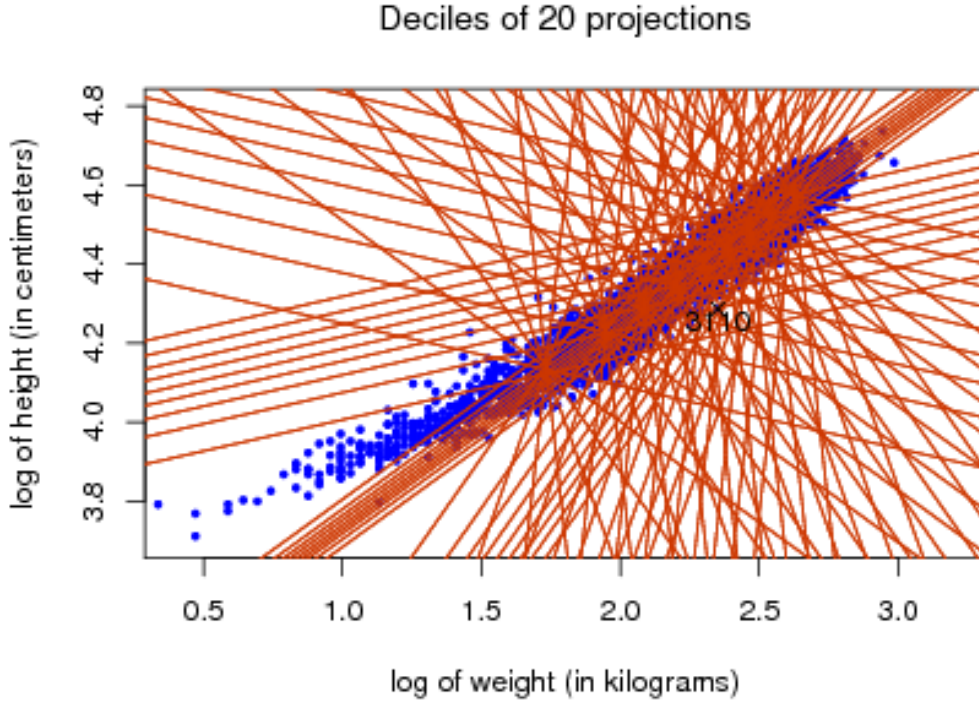


FIGURE 2. *The plot gets quickly overloaded if multiple directions and indexing probabilities are requested.*

It is formulated in terms of the Pompeiu-Hausdorff distance—the terminology we follow here is that of Rockafellar and Wets (1998).

**Theorem 2.** *Suppose that  $X$  is a random vector with distribution  $P$ . If the support  $P$  is bounded, then  $\mathcal{Q}(p, s, X)$  is a continuous function of  $s$ , for every  $p \in (0, 1)$ .*

*The same holds true when the support of  $X$  is contiguous; moreover, if a sequence of random vectors  $X_n$  converges almost surely to  $X$ , and  $s_n \rightarrow s$ , then  $\mathcal{Q}(p, s_n, X_n)$  converges to  $\mathcal{Q}(p, s, X)$  in the Pompeiu-Hausdorff distance, for every  $p \in (0, 1)$ .*

The theorem is formulated slightly more generally, to allow for alternative quantile versions, and to facilitate later asymptotic considerations. Using the theorem with  $X_n = X$  shows that the continuity of directional quantiles holds for all empirical, and many population distributions.

**5.3. Quantile biplots and directional quantile information.** Quantile biplots allow for faithful and relatively straightforward retrieval of the directional quantile information. Whenever a directional quantile line is sought, it is sufficient to find the intersection of the  $p$ -th contour with the halfline emanating from the selected origin

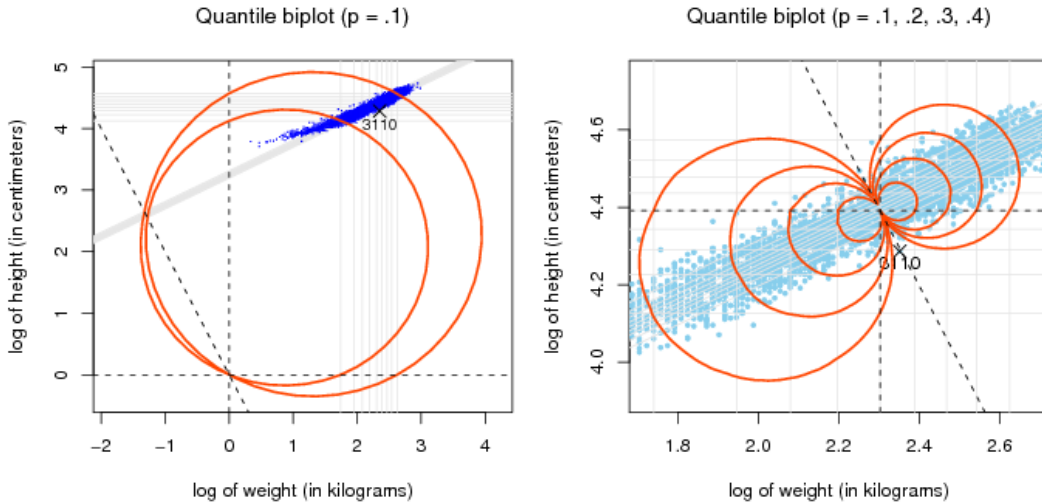


FIGURE 3. *Quantile biplots allow for faithful and relatively straightforward retrieval of the directional quantile information, but the counterintuitive character of contours, their dependence on the coordinate system, and certain other features (tendency to self-intersection and “mozzarella” shape) are rather disturbing.*

in the given direction  $s$  starting in the selected origin. The line passing through the intersection and perpendicular to this halfline is then the desired directional quantile line. The reader can check in Figure 3 how this works for the coordinate directions, and also for the “*BMI* direction”  $s \propto (-1/2, 1)$ .

However, otherwise quantile biplots exhibit several disturbing features. One of them, as revealed in the search for the origin of a quantile biplot, is the lack of any equivariance—even with respect to an operation as simple as a shift. For plotting, the equivariance with respect to translations and coordinatewise rescaling is something like a minimal requirement—otherwise the automatic rescaling implemented in typical graphical routines may easily distort the plotted content.

Overall, quantile biplot contours appear rather counterintuitive, and their tendency to self-intersections and “mozzarella” shapes, as in Figure 3, probably will not win them too many friends. It seems that the question is not how to plot directional quantiles, but how to successfully incorporate this information into the plot of the data. Directional quantile lines then appear still better than anything else—the only problem is to reduce the overload caused by their straightforward plotting.

## 6. NORMAL CONTOURS

**6.1. Multivariate quantiles via normal distribution.** Despite our lack of conviction in any need of a multivariate generalization of quantiles, let us, exclusively for this section, imagine a situation that we would be forced to furnish one. Under such an urgency, a suggestion of a classical statistical trainee would be to fit a normal distribution—in the faith that this distribution often captures the essence of the data, and in the wisdom that it is the most promising analytic form for the subsequent mathematical treatment. Once the decision is made, it remains only to call the contours of the fitted normal distribution “bivariate quantiles”—as symptomatically done by Evans (1982).

A technical question demanding clarification is that of indexing: which particular contours should correspond to which  $p$ ? As already discussed, the contours indexed by  $p$  and  $(1-p)$  are bound coincide—normal distribution is continuous and supported by the whole plane. Nonetheless, we still need to assign contours to  $p$  in  $(0, 1/2]$ , and there are essentially two ways of doing that.

**6.2. Indexing by the enclosed mass.** Indexing by the enclosed mass extrapolates the univariate fact that the  $p$ -th and  $(1-p)$ -th quantiles together leave  $2p$  of the distribution mass outside their convex hull. For example, the contours corresponding

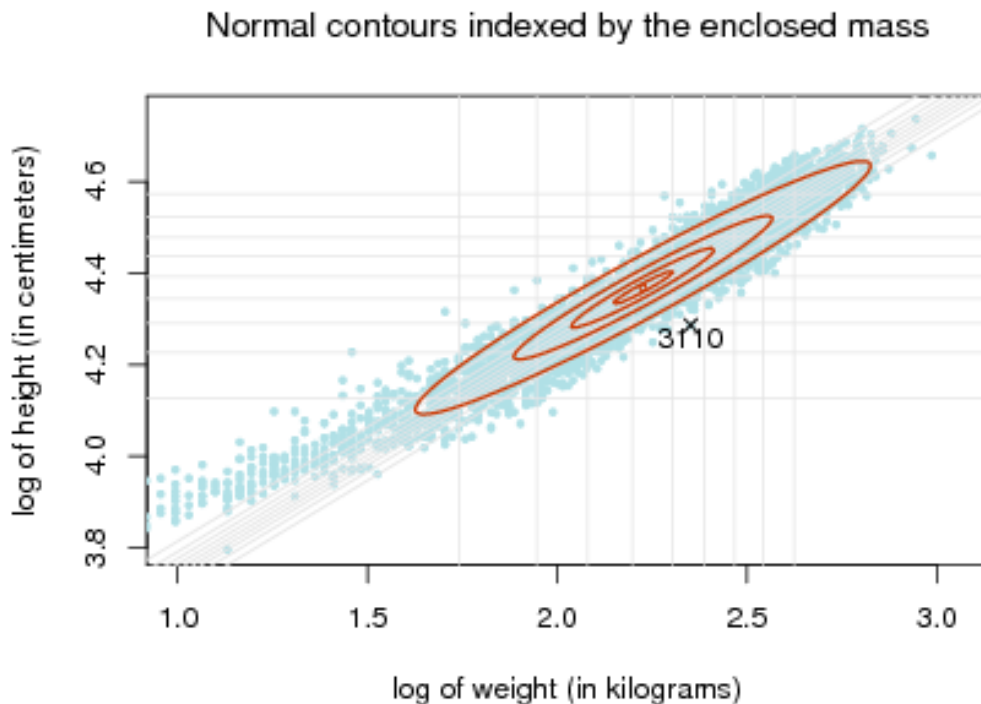


FIGURE 4. *If indexed by the mass they enclose, the contours of the fitted normal distribution do not interact well with directional and marginal quantiles.*

to deciles are those enclosing 0.8, 0.6, 0.4, 0.2 of the mass of the fitted normal distribution, together with the contour consisting of the single point located at the mode. The actual numerical values can be determined by a transformation to the standard bivariate normal distribution and the subsequent use of the Rayleigh distribution.

The result can be seen in Figure 4. From the interpretational point of view, we are able to observe that the subject represented by the point 3110 lies in the outstanding 20% of the sample; however, this exceptionality is somewhat “generic”—expressed not only through the company of similar subjects with large weight given the height, but also by the company of those with small weight given the height, and of those with small height and weight altogether.

Note the striking discrepancy between the marginal quantiles and fitted normal contours. Of course, we do not expect the latter to match the former exactly—after all, their constructions follow different principles, and while normal distribution fits

the data perhaps not that badly, this fit is in no sense ideal. Nevertheless, there would be no match even if normal distribution would be a perfect fit.

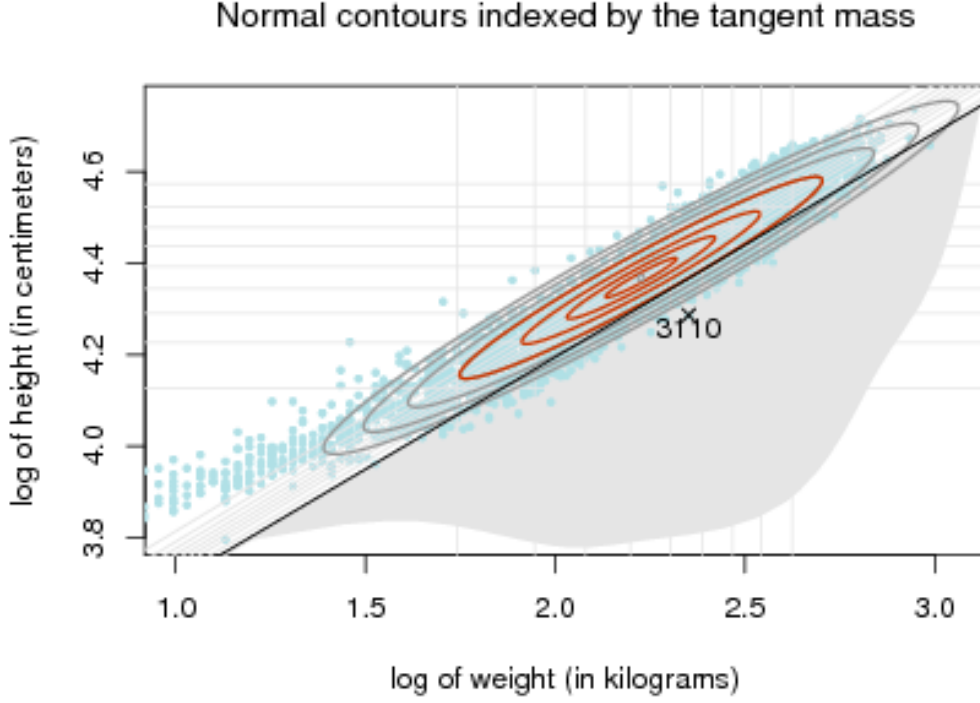


FIGURE 5. *If indexed by the tangent mass, the contours of the fitted normal distribution theoretically match projected quantiles. The half-plane tangent to the contour and passing through the point contains exactly  $p$  of the mass of the fitted multivariate normal distribution. Fitting normal contours has all virtues of the ideal, if the data “follow normal distribution”; here such compatibility appears to occur in the central part of the data, but not that much in tail areas, as demonstrated by the additional contours.*

**6.3. Indexing by the tangent mass.** An alternative way of indexing is that by the tangent mass, extrapolating the univariate fact that  $p$ -th and  $(1-p)$ -th quantiles mark the boundaries of the halfspaces containing exactly  $p$  of the distribution mass. For the standard normal distribution, the contour corresponding to  $p$  is that matching the univariate quantiles indexed by  $p$  and  $(1-p)$  when projected on the coordinate

axes, and can be found by transforming to the standard form and the subsequent inverse transformation.

The contours constructed in this way can be seen in Figure 5. Grey shading illustrated the following property: given any point of the contour, the halfplane passing through the point and tangent to the contour contains exactly  $p$  of the mass of the fitted multivariate normal distribution. The 10% extremality of 3110 can be thus interpreted not as “generic” now, but “substantial”: it is given by the company of subjects with similar nature, those with large weight given the height. Note that the boundary of the greyed halfspace is almost identical with the line indicating the (0.9)-th quantile of the *BMI*; hence the picture shows that in this case, the extremality of 3110 may be interpreted in terms of *BMI*.

Contrary to the indexing by the enclosed mass, indexing by the tangent mass interacts well with marginal and directional quantiles. Even if the match in Figure 5 is not perfect—as is obviously not the normal fit—it is better than in Figure 4. If the data follow normal distribution, then any directional quantile line can be retrieved as a line tangent to the contour in the given direction.

**6.4. Normal contours and directional quantile information.** Obviously, an approach based on fitting a parametric family of distributions is productive only if the data “follow that distribution”—that is, if the parametric family adapts well to the features present in the data. In a not-that-rare situation when this is not the case, an often attempted rescue is some ad hoc transformation to the canonical situation. Wei, Pere, Koenker and He (2005) demonstrate the pitfalls of such engineering, by showing that it may obscure features that can be unveiled by a more principled nonparametric methodology.

Otherwise, the approach through fitting normal contours could be considered ideal: the contours are a very good summary of the projected quantiles. After all, methods based on normality constitute the core of classical statistics—and despite all dissent, its *applied* core. It is thus not surprising that this way of thinking is usually the first (if not last) resort of practitioners; in the biometric context, the approaches to “reference contours” related somehow to normal distribution were pursued by Fatti, Senaoana and Thompson (1998) and Pere (2000).

Therefore, our—revised—ideal is to construct an alternative that would be non-parametric, but would extend the normal approach. We prefer indexing by the tangent mass, because it interacts much better with the directional and marginal quantile information than any other indexing convention. Finally, note that the

indexing calculations used transformations to and from the standard normal—for which the multivariate normal family is well suited, due to its *affine equivariance*—and thus brought up this important desideratum in quite an organic way.

## 7. DIRECTIONAL QUANTILE ENVELOPES AND HALFSPACE DEPTH

**7.1. The essence of our methodology.** To convey the information carried by directional quantile lines while avoiding the plotting overload, we propose to take, for fixed  $p$ , the inner envelope of the directional quantile lines. See Figure 6.

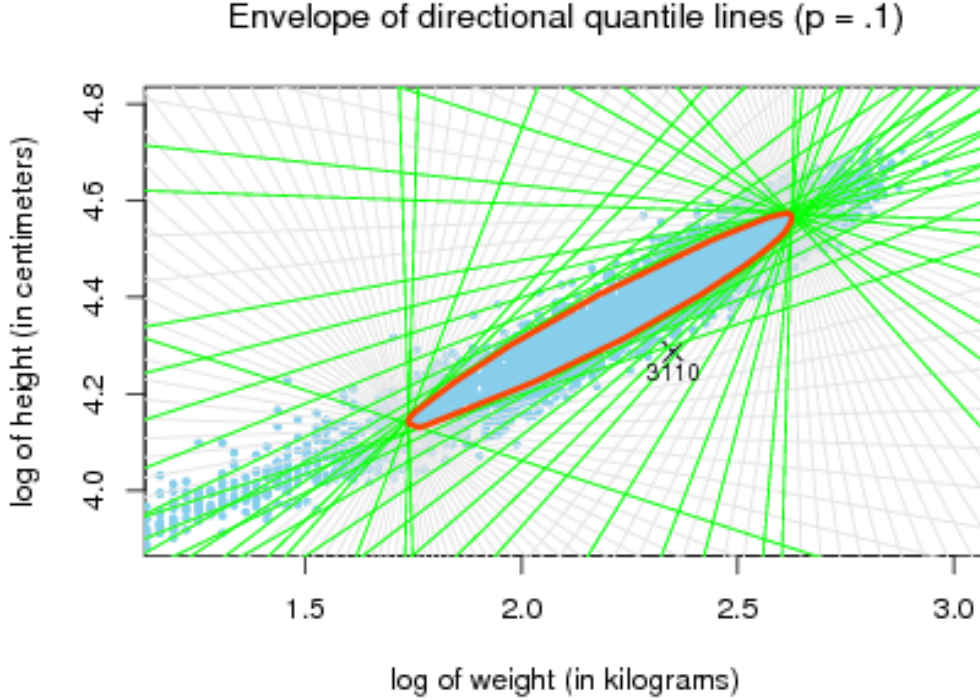


FIGURE 6. We propose to take, for fixed  $p$ , the inner envelope of the directional quantile lines.

**Definition 6.** Let  $X$  be a random vector with distribution  $P$ . For fixed  $p \in (0, 1/2]$ , the  $p$ -th *directional quantile envelope* generated by  $Q(p, s)$  is defined as the intersection,

$$D(p) = \bigcap_{s \in \mathbb{S}^{d-1}} H(s, Q(p, s)),$$

where  $H(s, q) = \{x: s^T x \geq q\}$  is the *supporting halfspace* determined by  $s \in \mathbb{S}^{d-1}$  and  $q \in \mathbb{R}$ . The notation  $D_A(p)$  will be used in case when the intersection is taken only over a subset  $A \subseteq \mathbb{S}^{d-1}$  of all possible directions; in the spirit of this notation,  $D(p) = D_{\mathbb{S}^{d-1}}(p)$ .

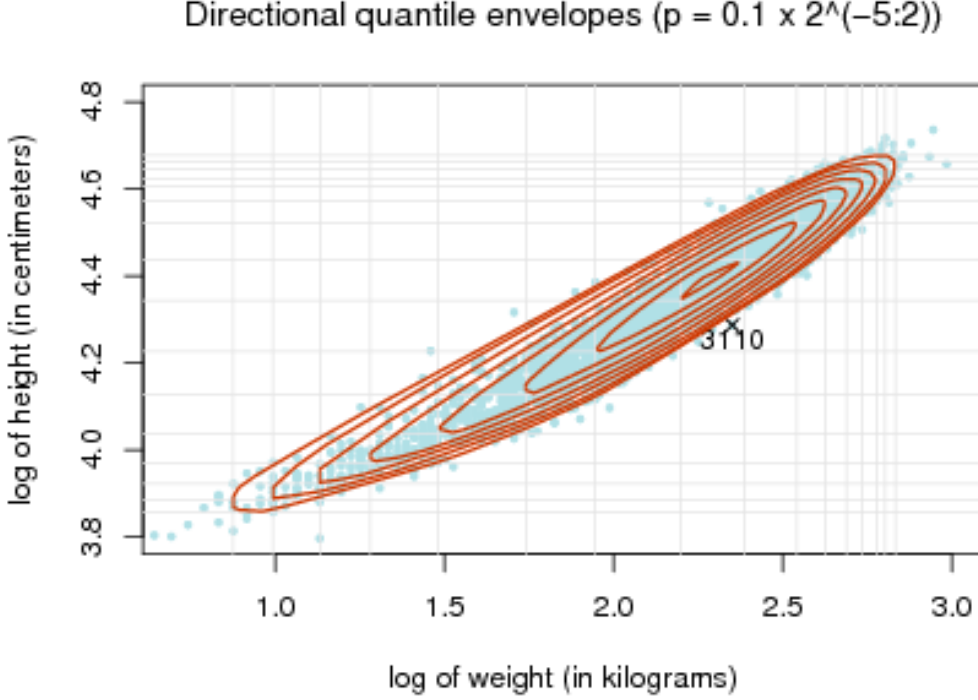


FIGURE 7. *Directional quantile envelopes for  $p = 2^i/10$ ,  $i = -5, \dots, 2$ . In the central part, the contours resemble those obtained by fitting normal distribution; in the tail area, they adapt more to the specific shape of the data. The plot can accommodate several  $p$  simultaneously, and the contours allow for relatively faithful and straightforward retrieval of the directional quantile information.*

In general, directional quantile envelopes are always bounded (if we are using a proper subset of directions to define  $D_A(p)$ , we usually want to ensure this by taking  $A$  not contained in any closed halfspace whose boundary contains the origin) and convex, being intersections of convex sets.

If we suppress the underlying directional quantile lines (still visible in Figure 6), we realize that the plot can accommodate several  $p$  simultaneously. Figure 7 shows

directional quantile envelopes for several  $p$  listed there. In the central part, the contours resemble those obtained by fitting normal distribution; in the tail area, they adapt more to the specific shape of the data. Theorem 5 below shows that the contours of normal distribution are actually directional quantile envelopes—hence we really do extend the approach from the previous sections. They may be empty; to learn more about this, we can use the following close relationship of directional quantile envelopes to a well-known data-analytic concept, depth.

**7.2. Directional quantile envelopes and halfspace depth.** The more specific denomination “halfspace depth” is sometimes used to distinguish the just defined depth from similar later notions. The commonly used name “Tukey depth” reflects that even though it was Hodges (1955) who first introduced it, Tukey (1975) proposed depth contours for plotting bivariate data, in a spirit close to ours. Other references are Donoho and Gasko (1992), Rousseeuw and Hubert (1999), Rousseeuw and Ruts (1999), and Mizera (2002).

**Definition 7.** Let  $P$  be a distribution in  $\mathbb{R}^d$ . The *depth*,  $d(x)$ , of a point  $x \in \mathbb{R}^d$ , is defined as  $\inf P(H)$ , where  $H$  runs over all closed halfspaces containing  $x$  (or, equivalently, over all closed halfspaces with  $x$  lying on their boundary).

**Theorem 3.** *For every  $p \in (0, 1/2]$ , the directional quantile envelope is equal to the upper level set of depth:  $D(p) = \{x: d(x) \geq p\}$ .*

Theorem 3 implies that directional quantile envelopes are nonempty for  $p \leq 1/(d+1)$ , in the two-dimensional case for  $p \leq 1/3$ , due to a basic result from depth theory known as a centerpoint theorem—see Donoho and Gasko (1992) or Mizera (2002).

The reason that why we refer to what are essentially depth contours as “directional quantile envelopes”, is the existence of various interpolated quantile versions—which we would prefer in practical use, in particular because they allow for constructing contours interpolating between various depth level sets. Also, while Theorem 3 is rigorously true only for the “inf” quantile version following Definition 1, most of the other theorems presented in this article are true also for other versions—the interested reader may find the discussion of these details at the end of every proof in the Appendix.

All interpolated versions of quantiles yield somewhat smaller envelopes; Rousseeuw and Ruts (1999) point out that this is also the case for the related notion of halfspace

trimmed contours of Massé and Theodorescu (1994). These subtle differences vanish in regular situations—for instance, for absolutely continuous distributions with positive densities.

**7.3. Directional quantile envelopes and support functions.** Except for the consequence of the centerpoint theorem, it turns out that there is not so much existing depth theory contributing to quantile tomography as directional quantile philosophy may shed light on depth. Mathematically, we work here with objects dual to convex sets, the support functions, as defined in convex analysis. In this way, we may continue the line of thought brought into statistics by Walther (1997a,b).

Recall that the support function of a convex set  $K$  is defined as  $\sigma_K(u) = \sup_{a \in A} a^T u$ . Every support function is *positively homogeneous*,  $\sigma_K(cu) = c\sigma_K(u)$  for every  $u$  and every  $c > 0$ , and *sublinear*,  $\sigma_K(u + v) \leq \sigma_K(u) + \sigma_K(v)$ . Conversely, every positively homogeneous and sublinear function is the support function of some convex set.

The equivariance of directional quantiles means that, for fixed  $p$ , the directional quantile function  $q(u) = Q(p, u^T X)$  is positively homogeneous; nevertheless, it is not hard to find an example when it is not sublinear. However, we can consider the maximal support function, that is, the maximal positively homogeneous and sublinear among those dominated by the directional quantile function. It turns out that the result is the support function of the directional quantile envelope. While this observation did not turn to be directly helpful in its technical aspect, it can be effectively used in the algorithm constructing bivariate directional quantile envelopes.

## 8. RECOVERY OF DIRECTIONAL QUANTILE INFORMATION

A question of paramount importance now is how far is it possible to recover the directional quantile lines from the directional quantile envelopes—what is the price for the compression of the directional quantile information.

Let  $e$  be a point lying on the boundary,  $\partial E$ , of a bounded convex set  $E \subset \mathbb{R}^d$ . A *tangent* of  $E$  at  $e$  is any hyperplane (line) containing  $e$  that has empty intersection with the interior of  $E$ . Such a line determines the corresponding *tangent halfspace*, the halfspace that has the tangent as its boundary and its interior does not contain any point of  $E$ . Let  $X$  be a random vector with the distribution  $P$ . The *maximal mass at a hyperplane* is defined as

$$\Delta(P) = \sup\{\mathbb{P}[s^T X = c] : s \in \mathbb{S}^{d-1}, c \in \mathbb{R}\}.$$

The following theorem is essential in interpreting directional quantile envelopes.

**Theorem 4.** *Let  $P$  be a distribution in  $\mathbb{R}^d$ , and let  $p \in (0, 1/2]$ . If  $H$  is a tangent halfspace of  $D(p)$ , then  $p \leq P(H) \leq 2p + \Delta(P)$ . Moreover,  $p \leq P(H) \leq p + \Delta(P)$ , if  $\partial H$  is the unique tangent of  $D(p)$  at some point from  $H \cap \partial D(p)$ ; in particular,  $P(H) = p$  if  $\Delta(P) = 0$ .*

*If  $A \in \mathbb{S}^{d-1}$  is a finite set of directions and  $H$  is a tangent halfspace of  $D_A(p)$ , then still  $P(H) \leq 2p + \Delta(P)$ , and  $p \leq P(H) \leq p + \Delta(P)$ , if  $\partial H$  is the unique tangent of  $D_A(p)$  at some point from  $H \cap \partial D(p)$ . In particular,  $P(H) = p$  if  $\Delta(P) = 0$ .*

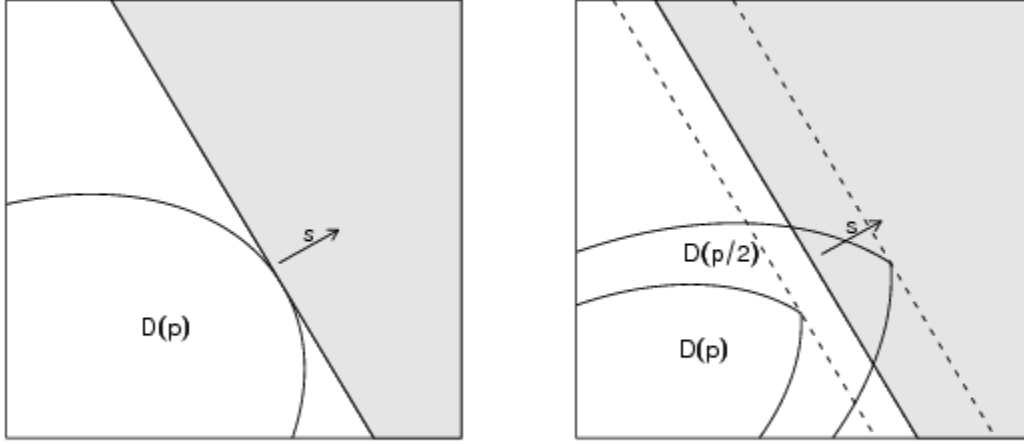


FIGURE 8. *Left panel: if the tangent line to the  $p$ -th directional quantile envelope is unique, then the tangential halfspace is the  $p$ -th directional quantile halfspace, in the given direction. Right panel: if the tangent line is nonunique, then this directional quantile halfspace lies between  $p$ -th and  $(p/2)$ -th directional quantile envelope.*

Theorem 4 provides a practical guideline how to recover the directional quantile information. The left panel of Figure 8 shows the situation when the tangent to the directional quantile envelope is unique. For a population distribution with  $\Delta(P) = 0$ , the user can uniquely identify the directional quantile in the direction perpendicular to the tangent. The visual determination of such a uniqueness may be slightly in the eye of beholder; if this is not desirable, the user may switch to a strictly finite-sample viewpoint, in which the directional quantile envelopes of empirical probability distributions are polygons and the uniquely identifiable directional quantile lines are those that contain a boundary segment of the polygon. Such a structure can be always discovered under appropriate magnification.

If the tangent is not unique, the situation shown in the right panel of Figure 8, then the exact identification of the directional quantile line is not possible; nevertheless, the inequality  $P(H) \leq 2p$  given by Theorem 4 allows at least for its approximate localization, especially when the plotted envelopes are so chosen that  $p$  follows a geometric progression with multiplier  $1/2$  (as in Figure 7; note that such choice gives approximately equispaced contours for normal distribution in the tail area).

A boundary point of a convex set that admits more than one tangent is called *rough* (singular). It is known—see Theorem 2.2.4 of Schneider (1993)—that such points are quite exceptional; in particular, for any closed convex set in  $\mathbb{R}^2$ , the set of rough points is at most countable. Convex, closed subsets of  $\mathbb{R}^d$  having no rough points are called *smooth*, consistently with the natural geometric perception of the boundary in this case. If  $D(p)$  is smooth, then the collection of its tangent halfspaces is in one-one correspondence with the collection of  $p$ -th directional quantile halfspaces, with the same boundaries, but in opposite directions.

Although the assumption of smoothness may sound optimistically mild, the examples in Rousseeuw and Ruts (1999) show that distributions with depth contours having a few rough points are not that uncommon. It may be argued that all these examples have somewhat contrived flavor, especially when the support of the distribution is some regular geometric figure. It is not impossible that typical population distributions have smooth depth contours—however, we were not able to find a suitable formal condition reinforcing this belief, beyond the somewhat restricted realm of elliptically-contoured distributions. Recall that the distribution is called *elliptic* if it can be transformed by an affine transformation to a circularly symmetric, rotationally-invariant distribution.

**Theorem 5.** *Every elliptic distribution has smooth directional quantile envelopes  $D(p)$ , for every  $p \in (0, 1/2)$ .*

In particular, this confirms the fact mentioned earlier: normal contours allow for the retrieval of *all* directional quantile lines.

**8.1. Characterization problem.** Even if the tangent line at a boundary point of a directional quantile envelope is nonunique, it does not necessarily mean that the information about certain directional quantiles is lost. Even if the directional quantile is not retrievable from the envelope *directly*, in a straightforward manner, it may be possible to reconstruct it from the totality of these envelopes. From the formal point of view, it means that the collection of directional quantile envelopes

for all  $p \in (0, 1/2]$  determines the distribution uniquely (and then, of course, all directional quantiles).

Surprisingly, this plausible property has not yet been rigorously proved in full generality. In the depth context, positive answers have been established for partial cases: depth functions uniquely characterize empirical (Struyf and Rousseeuw, 1999), and more generally atomic (Koshevoy, 2002) distributions, and also absolutely continuous distributions with compact support (Koshevoy, 2001). A small progress in this line is our following result regarding distributions with smooth depth contours. Note that these include, via Theorem 5, elliptic distributions, which may have unbounded support—hence the following theorem is not covered by that of Koshevoy (2001).

**Theorem 6.** *If the directional quantile envelopes  $D(p)$ , of a probability distribution  $P$  in  $\mathbb{R}^d$  with contiguous support, have smooth boundaries for every  $p \in (0, 1/2)$ , then there is no other probability distribution with the same directional quantile envelopes.*

## 9. ESTIMATION AND APPROXIMATION

**9.1. Beyond empirical distributions.** Since our discussion involved also a data example, the reader might get an impression that we have already proposed is a statistical methodology, that is, some algorithm(s) that can be used for processing the data. In fact, this has yet to be done; it is important to realize that our considerations so far were rather in a probabilistic than a statistical spirit.

Now, applying what was defined for general distributions to an empirical distribution indeed means some statistical advance, in the spirit of the principle of the approach called “naïve statistics” (Hájek and Vorlíčková, 1977) or “analogy” (Goldberger, 1968; Manski, 1988) or “plug-in” (Efron and Tibshirani, 1993): the result of the evaluation of a functional on the population distribution is estimated via the application of the same functional to the empirical distribution supported by the data. While this may be a way of obtaining satisfactory estimates (in fact, this is the exclusive approach considered in the depth literature so far) there are situations calling for more refined approaches.

Let us outline some general principles. We are interested in the *population quantile information*, that is, directional quantiles of some *population distribution*; we believe that our data come, in some sampling manner, from this distribution. To facilitate theoretical analysis of typical cases, it is often reasonable to posit some assumptions on this distribution; while a membership in a parametric family, or ellipticity may be considered too stringent, continuity assumptions are often acceptable.

The source from which the information is *estimated* are data. Our general strategy is, for fixed  $p$ , to estimate the directional quantiles  $Q(p, s)$  by  $\hat{Q}(p, s)$ , and then use these estimates to generate the estimated directional quantile envelope.

**9.2. Affine equivariance.** By Theorem 3, the estimated and population directional quantile envelopes are the level sets of depth applied to the empirical and population distributions, respectively. It is known that depth is affine invariant, and thus its level sets are affine equivariant.

**Definition 8.** An operator that assigns a point or a set  $T$  in  $\mathbb{R}^d$  to a collection of datapoints  $x_i \in \mathbb{R}^d$ , is called *affine equivariant*, if its value is  $BT + b$  when evaluated on the datapoints  $Bx_i + b$ , for any nonsingular matrix  $B$  and any  $b \in \mathbb{R}^d$ . (If  $T$  is a set, then the transformations are performed elementwise.)

In some other situations directional quantiles by some other means, for instance as a response of a quantile regression, then the affine equivariance of the resulting envelopes is not that clear. If the estimates exhibit some form of convergence to population depth contours, then one would have such an equivariance at least approximately; nevertheless, *exact* equivariance holds under mild assumptions on the directional quantile estimators.

**Theorem 7.** Suppose that directional quantile estimators  $\hat{Q}(p, s)$  are translation and scale equivariant, for all  $s \in \mathbb{S}^d$  and fixed  $p$ . Then the directional quantile envelope generated by these estimators is affine equivariant.

**9.3. Approximations.** There are several reasons to look at the effect of approximation on estimated envelopes. The numerical motivation stems from the fact that in practice we do not take all directions to construct the directional quantile envelope; therefore, what is constructed is rather an approximate envelope  $D_A(p)$ , and we are interested in the quality of this approximation. We know that  $D(p) \subseteq D_A(p)$ , and we believe that a decent collection  $A$  of directions that reasonably fill  $\mathbb{S}^{d-1}$  should make the approximation quite satisfactory. While the experimental evidence does not contradict this belief—for Figure 6 we used only 100 uniformly spaced directions, for Figure 7 we took 1009, and hardly any difference can be seen for  $p = 0.1$ —we would also like to have some theoretical support.

From the statistical point of view, our directional quantiles are usually not the “true”, but “estimated” ones. We believe, however, that this estimation possesses the usual consistent behavior—that is, we can show, in some customary probabilistic framework, that estimates become more and more precise, say, with growing sample

size. We are interested in whether this consistent behavior of individual directional quantiles translates to something analogous for their envelopes.

The following theorem, formulated in the general setting of supporting halfspaces, gives the essence of the relationship between directional quantiles and their envelopes.

**Theorem 8.** *Suppose that  $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$  is a sequence of closed sets with its union dense in a closed set  $A \subseteq \mathbb{S}^{d-1}$ , not contained in any closed halfspace whose boundary contains the origin. If for every sequence  $s_n \in A_n$  that converges to  $s \in A$ , the sequence  $q_n(s_n)$  converges to  $q(s)$ , then the sequence of sets  $\bigcap_{s \in A_n} H(s, q_n(s))$  converges to  $\bigcap_{s \in A} H(s, q(s))$  in the Pompeiu-Hausdorff distance—provided either the limit set is the closure of its interior, or it is a singleton and the sets in the sequence are nonempty.*

We may illustrate the use of this theorem on two instances. In the first, we take  $q_n(s) = q(s) = Q(p, s)$ ; the theorem then says that the successive approximations,  $D_{A_1}(p) \supseteq D_{A_2}(p) \supseteq \dots$ , approach  $D_A(p)$  in the Pompeiu-Hausdorff distance. Typically,  $A_n$  are finite, while  $A = \mathbb{S}^{d-1}$ ; the only requirements is that the directional quantiles  $Q(p, s)$  depend on  $s$  in a continuous way—for instance,  $P$  satisfies the assumptions of Theorem 2.

The second application furnishes a proof of consistency of  $\hat{D}_n(p)$  to  $D(p)$ , when  $\hat{D}_n(p)$  arise via applying the definition of directional quantile envelopes to empirical distributions that converge weakly almost surely to the sampled population distribution  $P$  (under suitable sampling scheme like independent sampling). Since the consistency of depth contours was discussed more thoroughly by He and Wang (1997), we consider this rather an example of the use of Theorem 8. The required assumptions are those of Theorem 2, continuous or bounded support of  $P$ , and the nondegeneracy of the limit  $D(p)$  (in general we cannot guarantee that  $\hat{D}_n(p)$  are nonempty). The Skorokhod representation yields random variables  $X_n$  converging almost surely to random variables  $X$ , such that the laws of  $X_n$  and  $X$  are the corresponding empirical distributions and  $P$ , respectively; Theorem 2 then implies the convergence assumption required by Theorem 8.

To obtain some idea about the magnitude of the approximation error, we can proceed as follows. For simplicity, we limit our scope to the two-dimensional setting. Let  $d \in \partial D$ . The directions of all tangents of a convex set  $D$  at  $d$  generate a convex cone,  $T_D(d)$ . Let  $c_D(d)$  be the maximal cosine between its two directions, the cosine

of the maximal angle between two extremal normalized directions in  $T_D(d)$ ,

$$c(d) = \sup \left\{ \frac{s^T t}{\|s\| \|t\|} : s, t \in T_D(d) \right\} = \sup \{s^T t : s, t \in T_D(d) \cap \mathbb{S}^{d-1}\}.$$

In fact, this cosine is the same as the maximal cosine of the directions in the normal cone  $N_D(d)$ ; see Rockafellar and Wets (1998), Chapter 6. We can see that  $c_D(d) \leq 1$ , the equality holding if and only if  $T_D(d)$  consists of single direction—when  $D$  has a unique tangent at  $d$ . Let

$$\kappa_D = \sup_{d \in \partial D} \sqrt{\frac{2}{1 + c(d)}},$$

the reciprocal of the cosine of the *half* of the maximal angle between directions in the tangent cone. Apparently,  $\kappa_D \geq 1$ , the equality holding true for smooth  $D$ . On the other hand,  $\kappa_D$  can be equal to  $+\infty$  for the degenerate  $D$ , the sets with empty interior.

**Theorem 9.** *Let  $A \subseteq \mathbb{S}^1$  be a set of directions, and let  $\hat{q}(s)$  and  $q(s)$  be two functions on  $A$ . Suppose that both  $\hat{D} = \bigcap_{s \in A} H(s, \hat{q}(s))$  and  $D = \bigcap_{s \in A} H(s, q(s))$  are nondegenerate; then both  $\kappa_{\hat{D}}$  and  $\kappa_D$  are finite and*

$$d(\hat{D}, D) \leq \max\{\kappa_{\hat{D}}, \kappa_D\} \sup_{s \in A} |\hat{q}(s) - q(s)|,$$

where  $d$  denotes the Pompeiu-Hausdorff distance.

## 10. SOME FURTHER DISCUSSION AND OTHER ASPECTS

**10.1. Indexing once again.** From the two ways of indexing discussed in Section 6, we prefer that by the tangent to that by the enclosed mass; we believe that its foremost advantage is the fact that it very naturally interacts with marginal (and projected) quantile information. We know that such judgment may be viewed as opportunistically adapted to the methodology we are trying to promote; indeed, directional quantile envelopes are naturally indexed by the tangent mass. The indexing decision may be a matter of individual choice—even authorities like Rousseeuw, Ruts and Tukey (1999) chose the “central box” in their bivariate generalization of boxplot to be the depth level set enclosing the half of the data—and we may represent rather a dissenting voice among those believing that indexing by the enclosed mass is that appropriate for “multivariate quantiles”, hypothetical objects being understood as some surrogates of confidence sets, with prescribed “coverage”.

Such desideratum of indexing by the enclosed mass motivated the proposal of Wei (2008), consisting roughly in fitting nonparametric quantile curves in polar coordinates to the data. The proposal of Wei (2008) focuses on conditional (with respect to the selected center) rather than directional quantiles; the result is dependent on the choice of the center—Wei (2008) uses the coordinate-wise median—and in general is not affine, nor orthogonally equivariant; nevertheless, it is equivariant with respect to translation and coordinate-wise rescaling (thanks to the preliminary normalization), so the minimal equivariance requirement for a plotting strategy, as discussed in Section 5.3, is met.

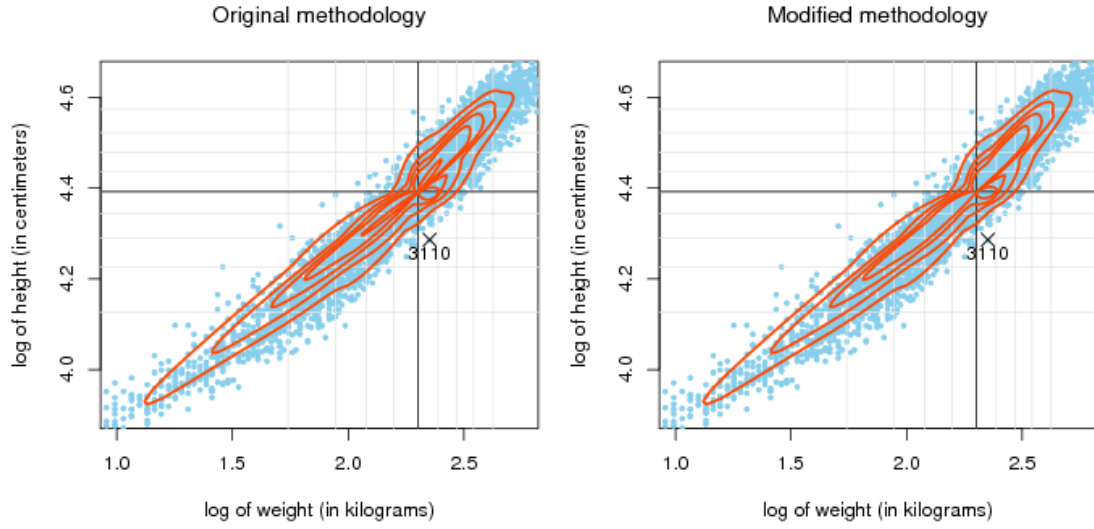


FIGURE 9. *It is not that impossible that some users may not like the shapes of contours generated by other methods — — —*

While the approach of Wei (2008) is conceptually capable of delivering interpretable contours, its practical application is plagued by its considerable dependence on the underlying nonparametric regression methodology. Especially in the hands of unskilled users (and the methodology has not been yet brought to an automated level excluding the adverse impact of those), the resulting contours may look like two specimens shown in Figure 9. Interestingly, the contours have some common virtues with our quantile biplots: a tendency to self-intersections, which is particularly strong when the prescription of Wei (2008) is followed faithfully, and conditional quantiles are fitted not on rays but on whole lines (left panel). This can be attenuated by slight modifications: by fitting conditional quantiles only on rays, and forcing

the fitted quantile regression to be nonnegative (right panel). If oversmoothed, the contours tend to extend outside of the data cloud, and follow “mozzarella shapes”; if undersmoothed, then they come out too rough, especially near the origin. For fitting the quantile curves we used the same R package `cobs` (Ng and Maechler, 2006) as did Wei (2008); best results were obtained when knots were placed in the quantiles of covariates—as recommended by the literature discussing regression splines, see, for instance Ruppert, Wand and Carroll (2003); the automatic knot selection seem not to improve the fit significantly (but slows the computation considerably).

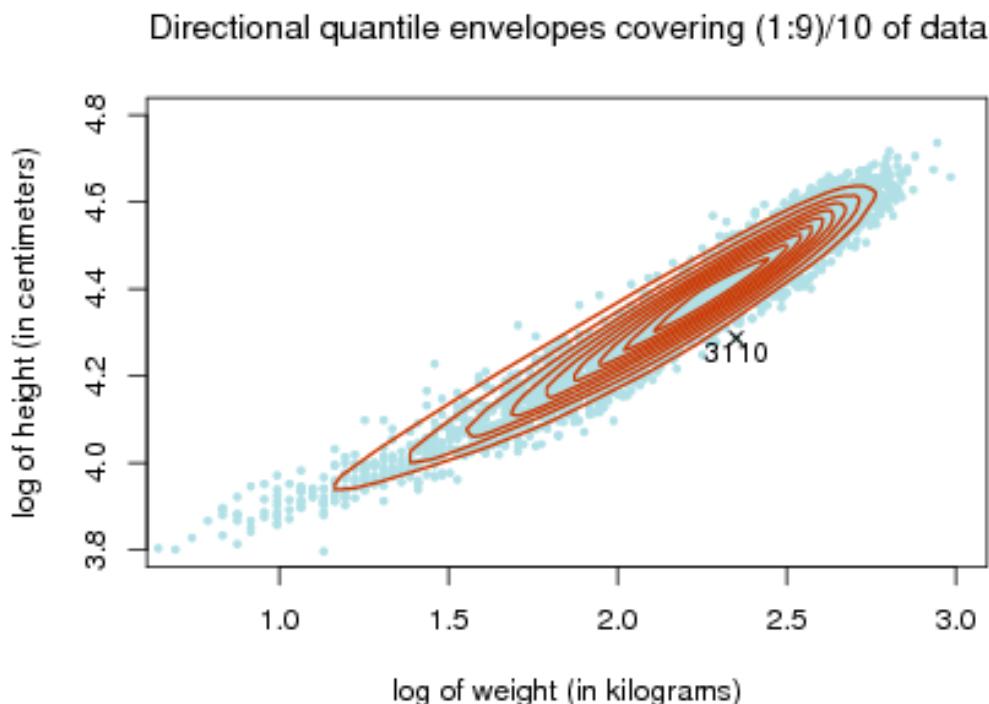


FIGURE 10. — in such a case, it is still possible to find directional quantile envelopes with desired coverage, despite of the different underlying philosophy.

It is not hard to imagine that some users would not feel that the contour like those shown in Figure 9 are those they really hoped to obtain. They might like more those from Figure 7 instead—would only they be indexed by the enclosed mass. In such a case, it is still possible to somewhat fulfill these desires by constructing directional quantile envelopes with prescribed coverage, employing a simple search. In fact,

this approach was adopted by Rousseeuw, Ruts and Tukey (1999) for their bivariate boxplot; indeed, the search is most easily accomplished in the location setting (in the absence of covariates), when depth levels sets are relatively quick to compute, and the search can be slightly sped up using the number of enclosed datapoints as an interpolating covariate. Nevertheless, such a strategy is not infeasible even in more sophisticated situations—for instance, in the quantile regression context—because finding the number of enclosed points does not require the actual construction of the envelope and thus is algorithmically quick.

Despite these possibilities, we do not advocate this way of indexing, but rather once again stress the strong interpretational appeal of indexing by the tangent mass. Even when realizing that the desideratum of “coverage” is codified in the daily standards of certain disciplines, we would still rather appeal to common sense—whether certain dogma cannot be changed. From this perspective, we view the approach of Wei (2008) rather complementary than alternative to ours. Despite all criticism, it conceptually addresses certain interpretational objectives—unlike various approaches based on minimum aggregate distances reviewed by Serfling (2002). While minimizing the total sum of the distances to the datapoints may have some potential in elucidating the “spatial” median, as a potential central point of the data, this approach does not convey any apparent statistical meaning when applied to quantiles.

**10.2. Estimation of the median.** The situation may change when we want to supplement the directional quantile information by a conforming estimate of the median. As follows from the properties of depth level sets discussed in Section 7.2, there is no guarantee that the directional quantile envelope for  $p = 1/2$ , and even for slightly smaller values of  $p$ , is nonempty. The approach developed in the depth literature (Donoho and Gasko, 1992) is that of the *Tukey median*: take the maximal depth level set (the nonempty level set with maximal  $p$ ), or a suitable point from it. As demonstrated by Rousseeuw, Ruts and Tukey (1999), this is again feasible in the location setting via a simple search strategy. In the presence of covariates, pursuing the conditional Tukey median may be not that easy—albeit not completely impossible; the main difficulty is that the maximal depth may vary with the covariate, which apart from algorithmic difficulties creates also certain conceptual puzzles.

The approaches based on minimal aggregate distance may seem to have an edge here; however, it should be reminded that they are usually not affine equivariant—unlike the least-squares algorithm for multivariate regression. While Koenker and Portnoy (1990) raise a question whether such a failure is “a mere peccadillo or a

mortal sin”, we would believe that once affine equivariance is forfeited, the most appealing solution, both from the conceptual and algorithmic aspects, is the intersection of coordinate-wise medians (or median regressions).

Since more suggestions pertinent to this problem are or will be provided by the literature on robust multivariate regressions, we believe we may stop discussing this topic here. We only conclude that our focus is not that much on the median rather than on quantiles; even if we want a median, then not just arbitrary, but one compatible with our directional quantile philosophy.

**10.3. Higher dimensions.** While the theoretical concepts expounded here extend in a more or less straightforward manner also to higher dimensions, we have to admit that computational complexity with growing dimension rapidly becomes prohibitive. However, given the inherent two-dimensionality of the plotting universe, it is dimension 2 where the proposed methodology is likely to be used—and there the algorithms work well. In fact, constructing a directional quantile envelope in its entirety is a task lacking a practical sense in the higher-dimensional context, where one would rather seeks local approaches: for instance, how much, in the tangent sense, is a given data unit exceptional? Some ideas in this direction have been outlined by Salibian-Barrera and Zamar (2006); much more remains to be done.

**10.4. Other properties.** The closer investigation of the process of forming the directional quantile envelope reveals that not all directional quantile lines have to be active in forming the exact or approximate envelope; this can be seen in Figure 6, where the active lines are the highlighted ones. If the inactive lines are omitted, the envelope remains the same, and to construct the approximate envelope, we actually need to do this elimination. An algorithm that accomplishes this with  $O(n)$  complexity for the approximate directional quantile envelopes with  $n$  directions will be described in subsequent work.

A side product of our investigations is the following guideline for selecting the origin for plotting quantiles in quantile biplots—which may be perhaps of some interest, should anybody find this way of plotting appealing. It turns out that such an origin should be as deep as possible; the best available one is thus located in the Tukey median, guaranteeing non-intersecting quantile lines for all  $p$  not exceeding its depth.

**Theorem 10.** *The curve formed by  $Q(p, s)$  for fixed  $p$  and revolving  $s$  does not intersect itself if the origin of the coordinate system has depth greater than  $p$ .*

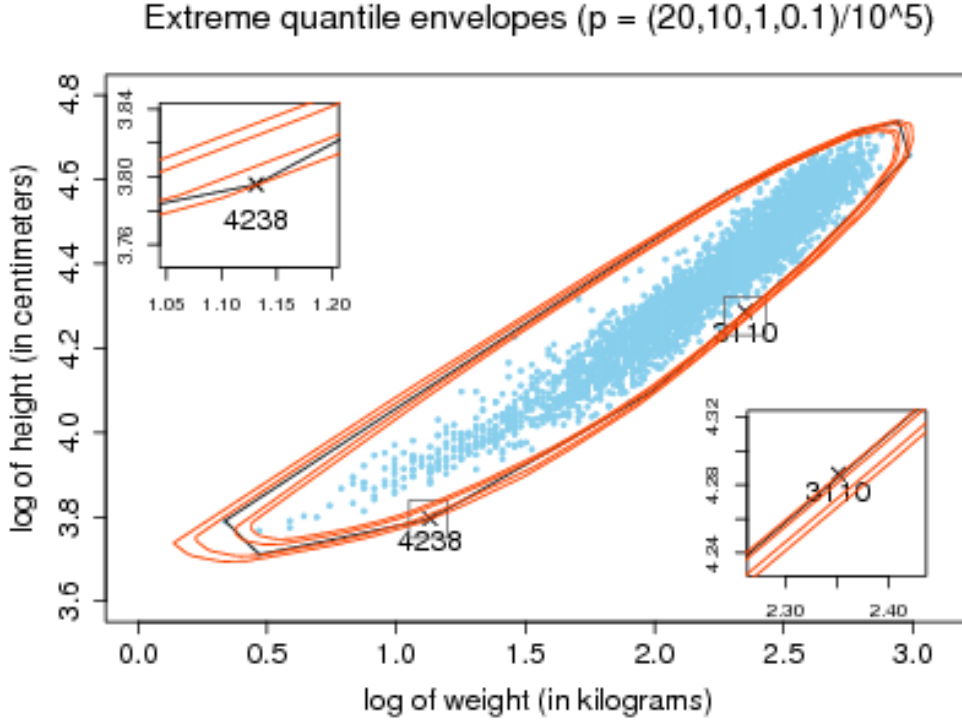


FIGURE 11. *Directional extremal quantiles, derived from the corresponding univariate analogs, and the convex hull, the empirical extremal quantile.*

## 11. DIRECTIONAL QUANTILE ENVELOPES IN VARIOUS STATISTICAL CONTEXTS

### 11.1. Simple location setting: no covariates, straightforward estimation.

We just remind the reader that we already demonstrated the use of our methodology above in the simple location setting, when there are no covariates and the estimation is performed via the application of the quantile operators to empirical distributions. Relevant illustrations are Figures 7 and 9; both estimate directional quantile envelopes by evaluating them for empirical distributions, and differ only in indexing convention.

**11.2. Extreme quantiles.** The case of extreme quantiles is the one where the need for other than empirical estimators of population quantiles is demonstrated very noticeably. If, say, 100 observations are available, then their maximum, the  $p$ -th

empirical quantile for any  $p > 0.99$ , may not be found satisfactory for estimating a threshold with exceedance probability less than, say, 0.001.

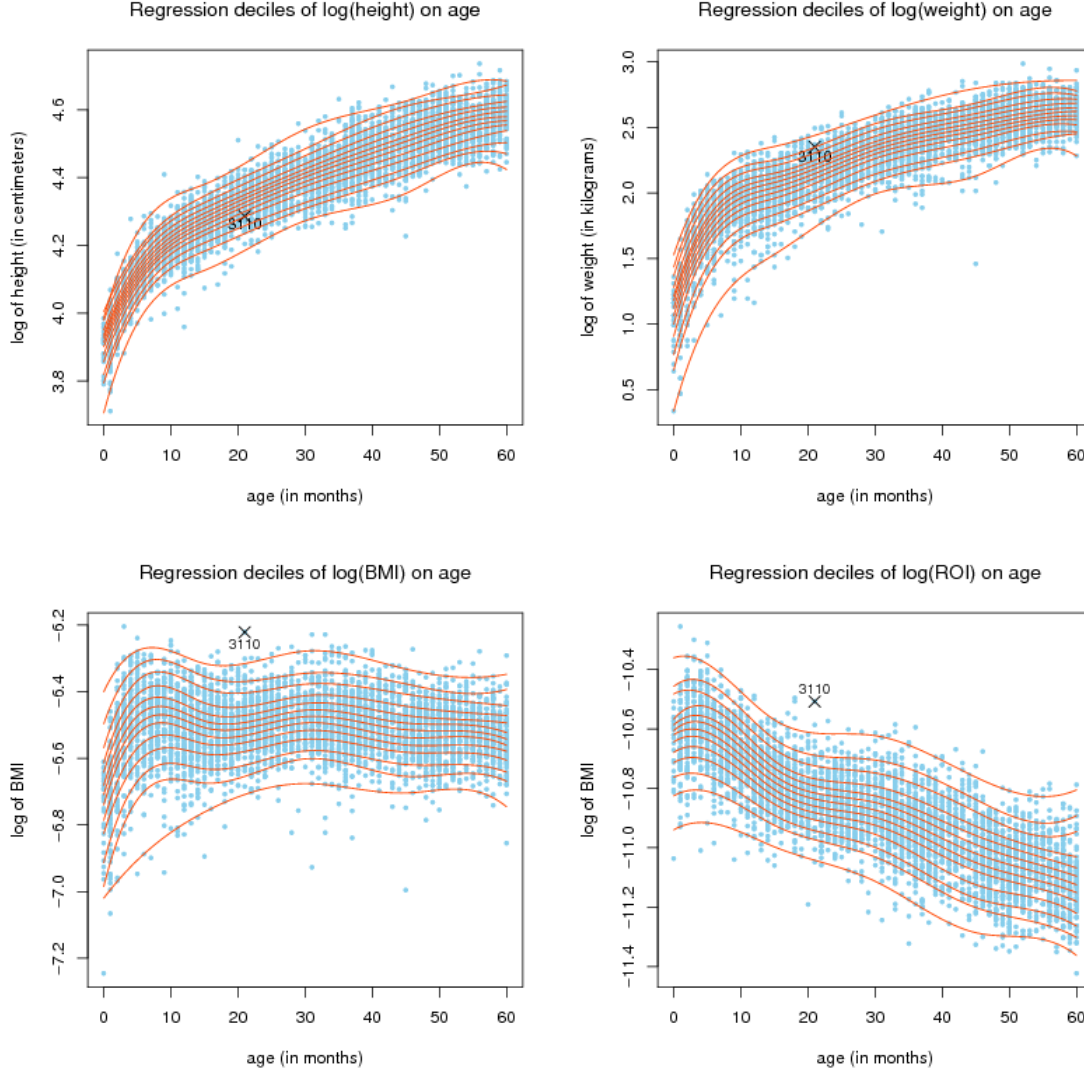


FIGURE 12. The “growth charts”, quantiles of various linear combinations of the primary variables, regressed on the covariate, age.

Various approaches to deal with this situation can be found in the books of Beirlant, Goegebeur, Teugels and Segers (2004), Reiss and Thomas (2007), Resnick (2007), and the references given there. We do not have a particular preference

for any of the methods proposed in that literature, and, being focused on the use of quantiles for multivariate data, we rather opportunistically chose from those with more nonparametric flavor the one we found implemented as the R package `evir` (McNeil and Stephens, 2007). That is, other approaches to extreme quantile estimation could and should be considered as well—as soon as they are implemented, they will fit our directional quantile scheme in the same vein.

The result can be seen in Figure 11. The estimated extreme quantiles, for  $p = 10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ , and  $2 \times 10^{-4}$  are confronted with the convex hull of the data, the empirical estimate for any  $p \leq (2.33)10^{-4}$ . The plot seems to provide some information about the extent of extremality of the points labeled by 3110 and 4238; a closer inspection reveals that 3110 lies on the  $(2 \times 10^{-3})$ -th directional quantile envelope, while 4238 on the  $(10^{-6})$ -th one. The real worth of this information is closely related to the same question regarding estimated extreme quantiles in the univariate case; nevertheless, if the pertinent discussion concludes favorably for some univariate alternative, then our methodology provides its viable multivariate extension.

**11.3. Quantile regression.** Figure 12 shows the deciles of several projections of the vector response, consisting of the logarithm of weight and height, regressed on the covariate, which is the age in months. While such “growth charts” are facilitating a lot of useful insights, the user may like to confront them with a directional perspective—in a related covariate-dependent context.

Such a desire stumbles upon the inevitable fact that our graphical universe is two-dimensional; animations and interactive graphics are certainly possible, but in traditional setting we can merely choose to plot directional quantiles for some fixed value(s) of the covariate—as in Figure 13, which shows the predicted envelopes for three values of the age (selected so that the resulting envelopes do not overplot, rather than pursuing any other objective). The highlighted datapoints represent the subjects with the particular age. If we computed directional quantile envelopes from these points separately, the resulting contours would be rougher, and would vary from one value of age to another; the contours presented in Figure 13 borrow strength from other ages, constructing quantile envelopes from a number of quantile regressions like those seen in Figure 12.

Once again, our focus here is on how quantile regression blends into directional quantile philosophy; hence our rendering of nonparametric quantile regression avoided rather than explored potential challenges. In view of Theorem 7, the only essential property is whether the estimates are translation (regression) and scale equivariant,

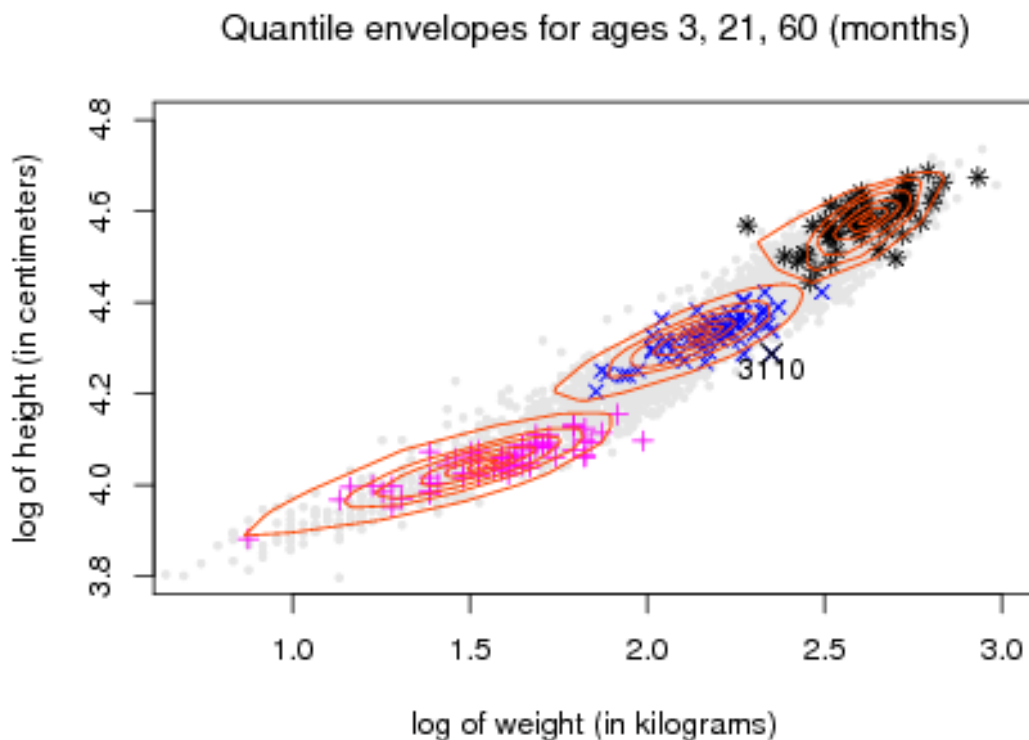


FIGURE 13. *Imagine the animation in which the directional quantile envelopes slowly ascend upward along the data cloud, demonstrating the dependence on the growing covariate, age.*

to yield affine equivariant envelopes. For various aspects of quantile regression, we refer to Koenker (2005) and references there. As Wei, Pere, Koenker and He (2005), we accomplished the fits in Figures 12 and 13 by regression splines, using the automated knot selection furnished by the R package `splines` (R Development Core Team, 2007), and fitting quantile regressions by the R package `quantreg` (Koenker, 2007). The smoothing parameter was selected by eyeballing the plots included in Figure 12, and then adopting a universal smoothing parameter for all directions in Figure 13. We are aware that while this may work well in certain situations (as it did in ours), one can easily imagine data exhibiting more signal-to-noise in one compared to other directions—then this fact should be reflected in variable smoothing

parameters. We hope to address this problem in future research, as well as explore the possibility of using alternative nonparametric quantile regressions strategies.

## 12. CONCLUSION

Directional quantile envelopes—essentially, depth contours—are a possible way to condense directional quantile information, the information carried by the quantiles of projections. In typical circumstances, they allow for relatively faithful and straightforward retrieval of the directional quantiles, and can be adapted to elaborate frameworks that require more sophisticated quantile estimation methods than evaluating quantiles for empirical distributions; these include estimation of extreme quantiles, and directional quantile regression. The resulting estimated quantile envelopes are affine equivariant under mild equivariance assumptions on the estimators of directional quantiles. The methodology offers straightforward probabilistic interpretations based on the concept of tangent mass.

We tried to clarify all the theoretical aspects of the proposed methodology as well as we could; we are aware of the fact that many important questions, as well as practical details, remain unanswered. We only hope that these gaps will be filled in further contributions to this theme.

## 13. ACKNOWLEDGEMENTS

We are indebted to Ying Wei for turning our attention to this problem, as well as for many insights contained in Wei (2008), and to Roger Koenker for valuable discussions. The directional approach to depth contours was pioneered in the unpublished master thesis of Benoît Laine, as reported by Koenker (2005)—in, however, quite significantly more complicated version fitting not directional quantiles, but directional quantile regressions. Another important forerunners of what is presented here are Salibian-Barrera and Zamar (2006).

## APPENDIX: PROOFS

**Proof of Theorem 1.** A direct consequence of Definition 1. For other quantile versions, the equivariance has to be checked individually—usually a straightforward task.

**Proof of Theorem 2.** Since quantile sets are bounded intervals, it is sufficient to prove the convergence of their endpoints to  $\inf \mathcal{Q}(p, s^T X) = \inf\{u: \mathbb{P}[s^T X \leq u] \geq p\}$  and  $\sup \mathcal{Q}(p, s^T X) = \sup\{u: \mathbb{P}[s^T X \geq u] \leq (1 - p)\}$ .

Suppose that the support of  $X$  is bounded. Let  $q = \inf \mathcal{Q}(p, s^T X)$ ; we have that  $\mathbb{P}[s^T X \leq q] \geq p$  and  $\mathbb{P}[s^T X \leq q - \varepsilon] < p$ . If the support of the distribution of  $X$  is bounded, we have  $\|X\| \leq M$  almost surely; by the Schwarz inequality,  $|(s - s_n)^T X| \leq M\|s - s_n\|$  and therefore

$$(2) \quad p \leq \mathbb{P}[s^T X \leq q] = \mathbb{P}[s_n^T X \leq q - (s - s_n)^T X] \leq \mathbb{P}[s_n^T X \leq q + M\|s - s_n\|],$$

which means that  $\inf \mathcal{Q}(p, s_n^T X) \leq q + M\|s - s_n\|$ . In a similar fashion, we obtain that  $\inf \mathcal{Q}(p, s_n^T X) \geq q - M\|s - s_n\| - \varepsilon$ , due to

$$(3) \quad \mathbb{P}[s_n^T X \leq q - M\|s - s_n\| - \varepsilon] \leq \mathbb{P}[s^T X \leq q - \varepsilon] < p.$$

Putting (2) and (3) together and letting  $\varepsilon \rightarrow 0$ , we obtain

$$q - M\|s - s_n\| \leq \inf \mathcal{Q}(p, s_n^T X) \leq q + M\|s - s_n\|,$$

and therefore  $\inf \mathcal{Q}(p, s_n^T X) \rightarrow \inf \mathcal{Q}(p, s^T X)$ , and thus also  $\mathcal{Q}(p, s_n, X_n) \rightarrow \mathcal{Q}(p, s, X)$ . The convergence of  $\sup \mathcal{Q}(p, s_n^T X) \rightarrow \sup \mathcal{Q}(p, s^T X)$  is proved analogously.

If the support of the distribution of  $X$  is contiguous, then all directional quantile sets in the limit are singletons. Pompeiu-Hausdorff convergence then follows from the “outer convergence” of quantile sets in the sense of Rockafellar and Wets (1998), see also Mizera and Volaufo (2002): any limit point,  $x$ , of any sequence  $x_n \in \mathcal{Q}(p, s_n, X_n)$  lies in  $\mathcal{Q}(p, s, X)$ . This can be easily seen in an elementary way, observing that  $x_n \in \mathcal{Q}(p, s_n, X_n)$  entails

$$p \leq \limsup_{n \rightarrow \infty} \mathbb{P}[s_n^T X_n \leq x_n] \leq \mathbb{P}[s^T X \leq x]$$

and

$$1 - p \leq \limsup_{n \rightarrow \infty} \mathbb{P}[s_n^T X_n \geq x_n] \leq \mathbb{P}[s^T X \geq x]$$

Since under the contiguous support assumption the quantiles are unique, this second part of the theorem holds true for every quantile version.

**Proof of Theorem 3.** If  $y \in D(p)$ , then  $y \in H(p, s)$  for every  $s \in \mathbb{S}^{d-1}$  and thus  $P(\{x: s^T x \geq s^T y\}) \geq p$  for all  $s \in \mathbb{S}^{d-1}$ ; therefore  $d(y) \geq p$ . Conversely, if  $d(y) \geq p$ , then for every  $s \in \mathbb{S}^{d-1}$  we have  $P(\{x: s^T x \geq s^T y\}) \geq p$ . It follows, in view of  $\inf$  in Definition 1, that  $s^T y \geq Q(p, s)$  and thus  $y \in H(p, s)$ . Hence  $y \in D(p)$ .

As already mentioned, this theorem is true only for the “inf” quantile version following Definition 1. Every other quantile version gives smaller envelopes.

**Proof of Theorem 4.** For  $A = \mathbb{S}^{d-1}$ , the inequality  $p \leq P(H)$  follows from the fact, established by Theorem 3, that the depth of every  $y \in D(p)$  is at least  $p$ .

Suppose that  $y \in \partial D_A(p)$ ; we claim that there is  $s \in A$  such that  $y \in \partial H(p, s)$ . Otherwise, there exists a  $\varepsilon > 0$  such that  $s^T y - Q(p, s) \geq \varepsilon$  for any  $s \in A$ . Then the ball centered at  $y$  with radius  $\varepsilon/2$  would belong to  $H(p, s)$  for any  $s \in A$  and thus belong to  $D_A(p)$  as well, that means,  $y$  is an interior point of  $D_A(p)$ , a contradiction.

Suppose now that  $\partial H$  is a unique tangent of  $D_A(p)$  at  $y$ , in the direction  $s \in A$ . That means that every other directional quantile hyperplane through  $y$ , in the direction  $t \in A$ ,  $t \neq s$ , has a common point with the interior of  $D_A(p)$ . An  $\varepsilon$ -ball centered at such a point is contained in  $H(p, t)$ , hence  $\partial H(p, t)$  is parallel to the tangent in distance more than  $\varepsilon > 0$ , and  $y$  lies in the interior of  $H(p, t)$ . Thus, for any direction  $t \neq s$ , the point  $y$  does not lie in  $\partial H(p, t)$ . Consequently,  $y \in \partial H(p, s)$ , and for  $H = \{x : s^T x \leq Q(p, s)\}$  we have  $P(H) \geq p$ .

On the other hand, if  $X$  is a random vector with distribution  $P$ , we have

$$P(H) = \mathbb{P}[s^T X \leq Q(p, s) - \varepsilon_n] + \mathbb{P}[Q(p, s) - \varepsilon_n \leq s^T X \leq Q(p, s)],$$

where  $\mathbb{P}[s^T X \leq Q(p, s) - \varepsilon_n] < p$  and  $\mathbb{P}[Q(p, s) - \varepsilon_n \leq s^T X \leq Q(p, s)] \leq \Delta(P)$  as  $\varepsilon_n \rightarrow 0$ . That means that  $P(H) \leq p + \Delta(P)$ .

It remains to prove the inequality  $P(H) \leq 2p + \Delta(P)$ . For simplicity, we assume that  $d = 2$ , the argument for  $d > 2$  being similar. Suppose  $u$  is one endpoint of the closed segment  $T_{D_A(p)}(y) \cap \mathbb{S}^1$ . Theorem 24.1 of Rockafellar (1996) implies that there is a sequence  $y_n \in \partial D_A(p)$  such that  $y_n \rightarrow y$  and  $y_n$  has a unique tangent with direction  $s_n$  such that  $s_n \rightarrow u$ . The convergences  $s_n^T X \rightarrow u^T X$  and  $s_n^T y_n \rightarrow u^T y$  imply that  $\liminf_{n \rightarrow \infty} \mathbb{P}[s_n^T X < s_n^T y_n] \geq \mathbb{P}[u^T X < u^T y]$ . Using the fact proved in the first part of the proof, we obtain  $s_n^T y_n = Q(p, s_n)$ , because  $y_n$  has unique tangent; this implies  $\mathbb{P}[s_n^T X < s_n^T y_n] < p$  and thus  $\mathbb{P}[u^T X < u^T y] \leq p$ . If  $v$  is the other endpoint of  $T_{D_A(p)}(y)$ , then we have  $\mathbb{P}[v^T X \leq v^T y] \leq p$ . Finally,  $t \in T_{D_A(p)}(y)$  implies that  $H \subseteq \{x : u^T x < u^T y\} \cap \{x : v^T x < v^T y\} \cup \{y\}$  and thus

$$P(H) \leq P(\{x : u^T x < u^T y\}) + P(\{x : v^T x < v^T y\}) + P(\{y\}) \leq 2p + \Delta(P).$$

**Proof of Theorem 5.** By rotational invariance, the directional quantile envelopes of any circularly symmetric distribution are circles; since elliptic distributions are those that can be transformed to the circular symmetric ones by an affine transformation, the theorem follows from their affine equivariance (and holds true for any quantile version).

**Proof of Theorem 6.** Without loss of generality, we may assume that the origin is the point with maximal depth. We claim that for any  $s \in \mathbb{S}^{d-1}$  and  $c \leq 0$ ,

$$P(\{x: s^T x \leq c\}) = p^*,$$

where  $p^* = \sup\{d(x) : s^T x = c\}$ . If  $p^* = 0$ , then the equation obviously holds true. When  $p^* > 0$ , we need to show that  $\{x: s^T x = c\}$  is tangent to  $D(p^*)$ , which is equivalent to  $\{x: s^T x = c\} \cap D(p^*)$  being a singleton as  $\partial D(p^*)$  is smooth. Otherwise, let  $y$  be one of the interior points in  $\{x: s^T x = c\} \cap D(p^*)$ ; then  $d(y) = p^*$ , that is,  $P(\{x: t^T x \leq t^T y\}) = p^*$  for some  $t \in \mathbb{S}^{d-1}$ . Meanwhile  $y$  is also an interior point of  $D(p^*)$ , and there exists a point  $z \in D(p^*) \cap \{x: t^T x \leq t^T y\}$  with  $t^T(z - y) < 0$ . Thus  $P(\{x: t^T x \leq t^T z\}) \geq p^*$ , which implies  $P(\{x: s^T z < t^T x < t^T y\}) = 0$ , a contradiction. Therefore,  $\{x: s^T x = c\}$  is tangent to  $D(p^*)$ ; and equivalently  $\{x: s^T x = c\} \cap D(p^*)$  is a singleton, denoted by  $y$ . Obviously  $d(y) = p^*$ . The fact that any non-tangent halfspace passing through  $y$  will contain interior points of  $D(p^*)$  implies that  $P(\{x: s^T x \leq c\}) = p^*$ . We have proved that for any  $s \in \mathbb{S}^{d-1}$ , the distribution of  $s^T X$  is uniquely determined by  $D(p)$ ; the theorem follows.

**Proof of Theorem 7.** Let  $B$  be a nonsingular matrix and  $b$  a vector. First, we verify that the transformation rule for the supporting halfspace of the directional quantile: for every  $s \in S$  and every  $p \in (0, 1)$ ,

$$(4) \quad H(B^*s/\|B^*s\|, Q(p, s, BX + b)) = BH(s, Q(p, s, X)) + b,$$

where  $B^* = (B^{-1})^T$ . Note that when  $B$  is orthogonal, then  $B^* = B$ , and when  $B$  is diagonal (more generally, symmetric), then  $B^* = B^{-1}$ . Indeed, the equation satisfied by  $x$  in  $BH(s, (Q, p, s, X))$ ,

$$s^T (B^{-1}x) \leq Q(p, s, X),$$

is equivalent to

$$((B^{-1})^T s)^T x = (B^*s)^T x \leq Q(p, s, X).$$

The norm of  $s$  is one, but not necessarily that of  $B^*s$ ; therefore, we divide both sides by  $\|B^*s\|$ ,

$$(5) \quad (1/\|B^*s\|)(B^*s)^T x \leq (1/\|B^*s\|)Q(p, s, X).$$

By the scale equivariance of the quantile operator, and by the relationship  $Q(p, s, AX) = Q(p, A^T s, X)$ , which follows directly from the definition, we obtain that the right-hand side of (5) is equal to

$$Q(p, s, X/\|B^*s\|) = Q(p, s/\|B^*s\|, X) = Q(p, B^*s/\|B^*s\|, BX).$$

Since the transformation  $BX + b$  is one-to-one, the transformed intersection of halfspaces is the intersection of transformed halfspaces. Therefore, the transformed directional quantile envelope is by (4)

$$\bigcap_{s \in \mathbb{S}^{d-1}} (BH(p, s, X) + b) = \bigcap_{s \in \mathbb{S}^{d-1}} H(p, B^*s/\|B^*s\|, BX + b).$$

The proof is concluded by observing that  $s \mapsto B^*s/\|B^*s\|$ , where  $B^* = (B^{-1})^\tau$ , is a one-to-one transformation of  $\mathbb{S}^{d-1}$  onto itself—as can be seen by the direct verification involving its inverse,  $t \mapsto B^\tau t/\|B^\tau t\|$ . The proof is the same for any quantile version.

**Proof of Theorem 8.** To prove convergence with respect to Pompeiu-Hausdorff distance, we exploit the following facts. First, the sequence  $\bigcap_{s \in A_n} H(s, q_n(s))$ , together with the limit  $\bigcap_{s \in A} H(s, q(s))$  is contained in a bounded set, starting from some  $n$ . This follows from the fact that sets  $A_n$  are approaching a dense set in  $A$ , and the latter is not contained in any halfspace whose boundary contains the origin; therefore this property is shared by  $A_n$  starting from some  $n$ , which means that

$$\bigcap_{s \in A_n} H(s, \inf_{k \geq n} q_k(s))$$

is the desired bounded set. For uniformly bounded sequences, the convergence in Pompeiu-Hausdorff distance follows from the convergence in Painlevé-Kuratowski sense; see Rockafellar and Wets (1998), 4.13. The latter means that a general sequence of sets  $K_n$  converges to  $K$  if (i) every limit point of any sequence  $x_n \in K_n$  lies in  $K$ ; (ii) every point from  $K$  is a limit of a sequence  $x_n \in K_n$ . See also Mizera and Volauk (2002).

For sequences of closed sets with “solid” limits, sets that are closures of their interior, the Painlevé-Kuratowski convergence follows from the “rough” convergence, defined by Lucchetti, Salinetti and Wets (1994) to require (i) together with (ii)’ every limit point of every sequence  $y_n \in (\text{int } K_n)^c$  is in  $(\text{int } K)^c$ . See also Lucchetti, Torre and Wets (1993). That is, one can replace outer and inner convergence requirement of Painlevé-Kuratowski definition by two outer convergences, the original one, and the other one for “closed complements”.

Suppose that  $y \in \text{int } K$ . Then  $y$  belongs to all but finitely many  $K_n$ ; otherwise, there would be a subsequence  $n_i$  such that  $y \in (\text{int } K_{n_i})^c$ , and by the modified version of (ii)’,  $y \in (\text{int } K)^c$ . Hence, every  $y$  from the relative interior of  $K$  is a limit of an (eventually constant) sequence  $y_n \in K_n$ . To obtain (ii) for every  $x \in K$ , consider a sequence  $y_k$  of points from (nonempty)  $\text{rint } K$  such that  $y_n \rightarrow y$ ; the

desired sequence  $x_n$  is then obtained by a “diagonal selection”: for every  $y_k$ , there is  $n_k$  such that  $y_k \in K_i$  for every  $i \geq k$ ; set  $x_n = y_k$  for every  $n_k \leq n < n_{k+1}$ .

Thus, it is sufficient to prove (i) and (ii)'. Suppose that  $x$  is a limit point of a sequence  $x_n \in \bigcap_{s \in A_n} H(s, q_n(s))$ . Then there is a subsequence such that  $s_n^T x_n \geq q_n(s_n)$  for every  $s_n \in A_n$ ; every  $s \in A$  is a limit of a sequence  $s_n \in A_n$ , therefore the assumptions of the theorem imply that  $s^T x \geq q(s)$ ; hence  $x \in \bigcap_{s \in A} H(s, q(s))$ . This proves (i). This proves theorem for the singleton case, since then the Painlevé-Kuratowski convergence is implied by (i) once the sets in the sequence are nonempty.

Suppose now that  $x$  is a limit point of a sequence  $x_n \in \left(\text{int} \bigcap_{s \in A_n} H(s, q_n(s))\right)^c$ , that is, a limit of some subsequence of  $x_n$ . Every such  $x_n$  satisfies  $s_n^T x_n \leq q_n(s_n)$  for some  $s_n \in A_n$ . By the compactness of  $A$ , there is  $s \in A$  that is a limit of a subsequence of  $s_n$ ; passing to the limit along the appropriate subsequences, we obtain that  $s^T x \leq q(s)$ , by the assumptions of the theorem. This means that  $x \in \left(\text{int} \bigcap_{s \in A} H(s, q(s))\right)^c$ .

**Proof of Theorem 9.** As  $\hat{D}$  and  $D$  are compact convex sets, we have  $d(\hat{D}, D) = d(\partial\hat{D}, \partial D)$ . Let  $\varepsilon = \sup_{s \in A} |\hat{q}(s) - q(s)|$ ; we will show that for any  $x \in \partial D$ ,  $d(x, \partial\hat{D}) \leq \kappa_D \varepsilon$ . Let  $\tilde{q}(s) = q(s) - \varepsilon$  and  $\tilde{D} = \bigcap_{s \in A} H(s, \tilde{q}(s))$ .

For simplicity, we assume that  $\tilde{D}$  is also nondegenerate. We have that  $\tilde{D} \subseteq \hat{D}$ , and also  $\tilde{D} \subseteq D$ , the latter set being congruent to  $\tilde{D}$ . If  $\kappa_D(x) > 1$ , then  $x$  is a vertex of  $D$ . Since  $d(x, \tilde{x}) = \kappa_D(x)\varepsilon$ , where  $\tilde{x}$  is the corresponding congruent vertex in  $\partial\tilde{D}$ , it follows that  $d(x, \partial\tilde{D}) \leq \kappa_D \varepsilon$ . When  $\kappa_D(x) = 1$ , then, by Theorem 24.1 of Rockafellar (1996), there exists a sequence  $x_n \neq x$ ,  $x_n \in \partial D$ , such that  $x_n \rightarrow x$  and  $s_n \rightarrow s$ ,  $\kappa_D(x_n) = 1$ , where  $s_n$  and  $s$  are the directions of the tangent lines passing through  $x_n$  and  $x$ , respectively. There are two possibilities.

If there is  $N$  such that  $s_n = s$  for any  $n > N$ , then there must be two points, denoted by  $y_1$  and  $y_2$ , in  $\partial H(s, q(s)) \cap \partial D$  such that  $\kappa_D(y_1) > 1$  and  $\kappa_D(y_2) > 1$ . That is,  $y_1$  and  $y_2$  are two vertices of  $D$  and there is no other vertex between  $y_1$  and  $y_2$  of  $D$ . Suppose that  $\tilde{y}_1$  and  $\tilde{y}_2$  are points congruent to them on  $\tilde{D}$ ; then  $\tilde{y}_1$  and  $\tilde{y}_2$  are two vertices of  $\tilde{D}$  and there is no other vertex between  $\tilde{y}_1$  and  $\tilde{y}_2$  of  $\tilde{D}$  as well. In other words, we have a trapezoid with vertices  $y_1$ ,  $y_2$ ,  $\tilde{y}_1$  and  $\tilde{y}_2$  and  $x$  lies on one of the bases. A simple geometric calculation then shows the existence of a point,  $y$ , lying on the base constructed by  $\tilde{y}_1$  and  $\tilde{y}_2$ , such that  $d(x, y) \leq \max\{\kappa_D(y_1), \kappa_D(y_2)\}\varepsilon$ , that is,  $d(x, \partial\tilde{D}) \leq \kappa_D \varepsilon$ .

Suppose that there is an infinite subsequence of  $s_n$  such that  $s_n \neq s$ . Let  $\partial\tilde{D}$  by  $\tilde{x}_n$  and  $\tilde{x}$  be the congruent counterparts of  $x_n$  and  $x$ , respectively; let  $s_n$  and  $s$  be the corresponding directions. Let  $y_n = \partial H(s_n, q(s_n)) \cap \partial H(s, q(s))$  and  $\tilde{y}_n = \partial\tilde{H}(s_n, q(s_n)) \cap \partial\tilde{H}(s, q(s))$ . We have that  $y_n \rightarrow x$ ,  $\tilde{y}_n \rightarrow \tilde{x}$ , and  $d(y_n, \tilde{y}_n) = \sqrt{2}\varepsilon/\sqrt{1+s_n^T s}$ . As  $d(y_n, \tilde{y}_n) \rightarrow d(x, \tilde{x})$  and  $\sqrt{2}\varepsilon/\sqrt{1+s_n^T s} \rightarrow \varepsilon$ , we arrive to  $d(x, \tilde{x}) = \varepsilon$ , which means  $d(x, \partial\tilde{D}) \leq \kappa_D \varepsilon$  again.

Taking into account that  $\tilde{D} \subseteq \hat{D}$ , we obtain that  $d(x, \partial\hat{D}) \leq \kappa_D \varepsilon$ , for any  $x \in \partial D$ . The theorem follows from this and the symmetric inequality,  $d(x, \partial D) \leq \kappa_{\hat{D}} \varepsilon$  holding true for any  $x \in \partial\hat{D}$ , which can be established in an analogous way.

**Proof of Theorem 10.** To show that the curve does not intersect itself, it is sufficient to prove that  $Q(p, s) \leq 0$  for any  $s \in \mathbb{S}^{d-1}$ . As  $p \leq d(0)$ , we have

$$P(\{x : s^T x \leq Q(p, s)\}) = p \leq d(0) \leq P(\{x : s^T x \leq 0\}).$$

Therefore we have  $Q(p, s) \geq 0$  for any  $s \in \mathbb{S}^{d-1}$ .

## REFERENCES

- ARONSON, J. K. (2001). Francis Galton and the invention of terms for quantiles. *Journal of Clinical Epidemiology* **54** 1191–1194.
- BEIRLANT, J., GOEGBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester (with contributions from D. De Waal and C. Ferro).
- DONOHU, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.
- EDGEWORTH, F. Y. (1886). Problems in probabilities. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 5th series* **22** 371–384.
- EDGEWORTH, F. Y. (1893). Exercises in the calculation of errors. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 5th series* **36** 98–111.
- EFRON, B. and TIBSHIRANI, R. J. (1993). *An introduction to bootstrap*. Chapman and Hall, New York.
- EUBANK, R. L. (1986). Quantiles. In *Encyclopedia of Statistical Sciences, Volume 7* (S. Kotz, N. L. Johnson and C. B. Read, eds.) 424–432. Wiley, New York.
- EVANS, M. (1982). Confidence bands for bivariate quantiles. *Commun. Statist.-Theor. Meth.* **11** 1465–1474.

- FATTI, L. P., SENAOANA, E. M. and THOMPSON, M. L. (1998). Bayesian updating in reference centile charts. *J. R. Statist. Soc. A* **161** 103–115.
- FROHNE, I. and HYNDMAN, R. J. (2004). `quantile`. A function in R starting from version 2.0.0, <http://www.r-project.org>.
- GALTON, F. (1888-1889). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London* **45** 135–145.
- GOLDBERGER, A. S. (1968). *Topics in regression analysis*. Macmillan, New York.
- HÁJEK, J. and VORLÍČKOVÁ, D. (1977). *Mathematical statistics*. SPN, Praha [in Czech].
- HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504.
- HODGES, J. L., JR (1955). A bivariate sign test. *Ann. Math. Statist.* **26** 523–527.
- HYNDMAN, R. J. and FAN, Y. (1996). Sample quantiles in statistical packages. *The American Statistician* **50** 361–365.
- KOENKER, R. (2005). *Quantile regression*. Cambridge University Press, Cambridge.
- KOENKER, R. (2007). `quantreg`: *Quantile Regression*. R package version 4.10, <http://www.r-project.org>.
- KOENKER, R. and PORTNOY, S. (1990). M estimation of multivariate regressions. *J. Amer. Statist. Assoc.* **85** 1060–1068.
- KOSHEVOY, G. A. (2001). Projections of lift zonoids, the Oja depth and Tukey depth. Preprint.
- KOSHEVOY, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multiv. Anal.* **83** 360–364.
- LUCCHETTI, R., TORRE, A. and WETS, R. J.-B. (1993). Uniform convergence of probability measures: topological criteria. *Canadian Mathematical Bulletin* **36** 197–208.
- LUCCHETTI, R., SALINETTI, G. and WETS, R. J.-B. (1994). Uniform convergence of probability measures: topological criteria. *Journal of Multivariate Analysis* **51** 252–264.
- MANSKI, C. F. (1988). *Analog estimation methods in econometrics*. Chapman and Hall, New York.
- MASSÉ, J.-C. and THEODORESCU, R. (1994). Halfplane trimming for bivariate distributions. *J. Multivariate Anal.* **48** 188–202.
- MCNEIL, A. and STEPHENS, A. (2007). `evir`: *Extreme Values in R*. S original (EVIS) by Alexander McNeil and R port by Alec Stephens, R package version 1.5, <http://www.maths.lancs.ac.uk/~stephena/>.

- MIZERA, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736.
- MIZERA, I. and VOLAUF, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.* **83** 365–388.
- NG, P. T. and MAECHLER, M. (2006). `cobs`: *COBS – Constrained B-splines (Sparse matrix based)*. R package version 1.1-3.5, <http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:cobs>.
- PARZEN, E. (2004). Quantile probability and statistical data modeling. *Statist. Sci.* **19** 652–662.
- PERE, A. (2000). Comparison of two methods for transforming height and weight to normality. *Annals of Human Biology* **27** 35–45.
- R DEVELOPMENT CORE TEAM (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, <http://www.R-project.org>.
- REISS, R.-D. and THOMAS, M. (2007). *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel.
- RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- ROCKAFELLAR, R. T. (1996). *Convex analysis*. Princeton University Press, Princeton.
- ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational analysis*. Springer-Verlag, Berlin.
- ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94** 388–402.
- ROUSSEEUW, P. J. and RUTS, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244.
- ROUSSEEUW, P. J., RUTS, I. and TUKEY, J. W. (1999). The bagplot: A bivariate boxplot. *Amer. Statist.* **53** 382–387.
- RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric regression*. Cambridge University Press, Cambridge.
- SALIBIAN-BARRERA, M. and ZAMAR, R. (2006). Discussion: Conditional growth charts (by Y. Wei and X. He). *Annals of Statistics* **34** 2113–2118.
- SCHNEIDER, R. (1993). *Convex bodies: The Brunn-Minkowski theory*. Cambridge University Press, Cambridge.
- SERFLING, R. (2002). Quantile functions for multivariate analysis: approaches and

- applications. *Statistica Neerlandica* **56** 214–232.
- SHORACK, G. R. (2000). *Probability for statisticians*. Springer-Verlag, New York.
- STRUYF, A. and ROUSSEEUW, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *J. of Multiv. Anal.* **69** 135–153.
- TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2 523–531. Canad. Math. Congress, Quebec.
- WALTHER, G. (1997a). Monte Carlo sampling in dual space for approximating the empirical halfspace distance. *Annals of Statistics* **25** 1926–1953.
- WALTHER, G. (1997b). On a conjecture concerning a theorem of Cramér. *J. Multiv. Analysis* **63** 313–319.
- WEI, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.* **103** 397–409.
- WEI, Y., PERE, A., KOENKER, R. and HE, X. (2005). Quantile regression methods for reference growth charts. *Statistics in medicine* **25** 1369–1382.

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES  
UNIVERSITY OF ALBERTA  
CAB 632, EDMONTON, ALBERTA, T6G 2G1 CANADA  
mizera@stat.ualberta.ca  
<http://www.stat.ualberta.ca/~mizera>